

# The Wishart planted ensemble: A tunably-rugged pairwise Ising model with a first-order phase transition

Firas Hamze,<sup>1</sup> Jack Raymond,<sup>1</sup> Christopher A. Pattison,<sup>2</sup> Katja Biswas,<sup>2</sup> and Helmut G. Katzgraber<sup>3,2,4</sup>

<sup>1</sup>*D-Wave Systems, Inc., 3033 Beta Avenue, Burnaby, British Columbia, V5G 4M9, Canada*

<sup>2</sup>*Department of Physics and Astronomy, Texas A&M University, College Station, Texas 77843-4242, USA*

<sup>3</sup>*Microsoft Quantum, Microsoft, Redmond, WA 98052, USA*

<sup>4</sup>*Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, New Mexico 87501 USA*

(Dated: June 4, 2019)

We propose the Wishart planted ensemble, a class of zero-field Ising models with tunable algorithmic hardness and specifiable (or planted) ground state. The problem class arises from a simple procedure for generating a family of random integer programming problems with specific statistical symmetry properties, but turns out to have intimate connections to a sign-inverted variant of the Hopfield model. The Hamiltonian contains only 2-spin interactions, with the coupler matrix following a type of Wishart distribution. The class exhibits a classical first-order phase transition in temperature. For some parameter settings the model has a locally-stable paramagnetic state, a feature which correlates strongly with difficulty in finding the ground state and suggests an extremely rugged energy landscape. We analytically probe the ensemble thermodynamic properties by deriving the Thouless-Anderson-Palmer equations and free energy and corroborate the results with a replica and annealed approximation analysis; extensive Monte Carlo simulations confirm our predictions of the first-order transition temperature. The class exhibits a wide variation in algorithmic hardness as a generation parameter is varied, with a pronounced easy-hard-easy profile and peak in solution time towering many orders of magnitude over that of the easy regimes. By deriving the ensemble-averaged energy distribution and taking into account finite-precision representation, we propose an analytical expression for the location of the hardness peak and show that at fixed precision, the number of constraints in the integer program must increase with system size to yield truly hard problems. The Wishart planted ensemble is interesting for its peculiar physical properties and provides a useful and analytically-transparent set of problems for benchmarking optimization algorithms.

## I. INTRODUCTION

The interface between physics and computational complexity has yielded fruitful insights over decades of research. Hard optimization problems—which are ubiquitous throughout the natural sciences and domains such as operations research—are of significant importance to humans and are widely believed to admit no efficient algorithms for their solution over all members of their class. It was recognized that such problems show analogous features to those found in statistical mechanical systems, for example the existence of algorithmic phase transitions [1, 2], under which typical problems show a dramatic increase in the difficulty faced by known exact and heuristic algorithms.

In some cases, insights from the physics of spin glasses and disordered systems have inspired remarkable new algorithms; for example Mézard, Parisi, and Zecchia [3] studied algorithmic hardness transitions in random 3-satisfiability (3-SAT) problems using tools from statis-

tical physics subsequently proposed survey propagation as a promising method for solving such problems. In addition, physics-based approaches have suggested ensembles of very hard problems; for example locked constraint satisfaction problems [4] owe their difficulty to the fragmentation of the solution space into widely-separated sets.

An important special class of hard problem ensembles are those whose solutions are known to the constructor; these are often known as optimization problems with planted solutions. Aside from their theoretical interest, such problems are noteworthy for several reasons. They may, for example, serve as candidates for cryptographic one-way functions, that is functions whose outputs are cheaply computable for any input but for which determining an input yielding a given output is hard. Furthermore, they serve as useful benchmark problems for evaluating heuristic or exact algorithms. In recent years, the need for such benchmarks has increased with the advent of physical devices implementing quantum annealing [5]

and related (e.g. [6]) algorithms. In such situations, it is desirable to not only have access to a set of problems of tunable hardness but to also be able to compare an algorithm’s performance with the correct answer.

A physics-based approach for generating hard 3-SAT problems with planted solutions was proposed by Barthel *et al.* [7]; more recently, Krzakala and Zdeborová presented a technique known as quiet planting [8] for devising graph  $q$ -coloring problems with known solutions whose properties are indistinguishable from those of a random ensemble. This concept can be generalized to a variety of sparse problems, and has a close connection to reconstruction on trees [9].

While the aforementioned techniques and analyses have yielded numerous elegant insights, they all share the undesirable property of considering problems that are structurally far-removed from contemporary physics-based optimization devices.  $K$ -satisfiability problems, for example, require energy functions to include terms whose value depends on groups of  $K$  variables; in all the previously-mentioned work over binary variables,  $K \geq 3$ . Realistic spin system models and optimization hardware on the other hand are typically restricted to pairwise interactions; emulating higher-order interactions on such systems can require tremendous overhead. In contrast,  $q$ -coloring problems directly correspond to the antiferromagnetic Potts model of statistical physics and are thus expressible in terms of 2-body interactions, but  $q$  must be larger than 2 to yield hard problems as 2-coloring instances can be solved in linear time. On devices natively encoding problems consisting of binary variables, this can be problematic. Various techniques have recently been published [10–13] for constructing planted Ising instances on sparse graph topologies, which in some cases appear to yield quite difficult problems [12], but in common with short-ranged disordered models in general, most of their known properties are inferred from numerical simulation and much remains unexplored about which features make them amenable as benchmarks for given algorithms.

In this paper, we propose simple randomized procedure for generating systems of binary-constrained integer programs with a known solution [14]. The system coefficients are generated according to a specific type of correlated multivariate Gaussian distribution. When translated to an effective Ising Hamiltonian, we obtain a novel type of disordered system with known ground state which we call the Wishart planted ensemble; the name is inspired by the distribution followed by the resultant random matrix of couplers.

The Hamiltonian includes interactions among all pairs of variables; this aspect makes it a less-than-perfect fit for testing on devices which implement short-range topologies. Unlike hard problem ensembles based on  $K$ -SAT or graph coloring however, the variable domains are binary-valued and the interactions are pairwise. Vitally, the ensemble displays an rich array of thermodynamic and computational properties of considerable relevance to both classical and quantum algorithms. Computationally, the problems can be tuned to range in difficulty from very easy to extremely difficult at quite modest system sizes. We emphasize that nearly all statements about problem “hardness” in this paper refer to the empirically-observed *typical-case* difficulty encountered by heuristics (and in all likelihood exact algorithms) and have no bearing on theoretical computer science questions concerned with *worst-case* difficulty.

A striking physical property of the Wishart ensemble is the existence of a first-order phase transition, a discontinuous jump in the free energy at some system-dependent critical temperature. A key parameter in the generation procedure is  $\alpha$ , which specifies the number-of-equations-to-number-of-variables ratio;  $\alpha$  is analogous to parameters such as the clause-to-variable ratio in the satisfiability problems and exerts critical influence on the physical and algorithmic complexity. By deriving the Thouless-Anderson-Palmer [15] equations for the ensemble with special care to account for the correlated couplers, we obtain the mean-field free energy, from which we find that at any finite  $\alpha$ , the internal energy jumps discontinuously from some excited value to that of the planted ground state energy at some  $\alpha$ -dependent critical temperature  $T_c$ . We verify this temperature and the nature of the transition with extensive parallel tempering Monte Carlo [16, 17] simulations, which reveal that the system converges to its asymptotic predicted properties quite rapidly. For large  $\alpha > 1$ , the magnitude of the discontinuity decreases monotonically and the thermodynamics smoothly change character towards a traditional second-order ferromagnetic transition. When  $\alpha < 1$ , the paramagnetic state, i.e., the set of all configurations having no correlation with the planted solution, is stable at any nonzero temperature. This feature signals difficulty for classical heuristic algorithms as it behaves as a deceptive dynamical “trap.” More specifically, following free energy gradients as done by methods like simulated annealing [18] will overwhelmingly lead to solutions far from the true optimum. Only by fortuitous initialization within the ground state basin of attraction will the problem be solved with high probability. It turns out that  $\alpha$

modulates the size of the ground state basin, with larger values increasing the probability of solution by lucky initialization. When  $\alpha \geq 1$ , on the other hand, the paramagnetic state becomes unstable for some temperature  $T_u < T_c$ ; at that point local algorithms can successfully find the solution by “rolling downhill” and hence such problems are typically easy. Remarkably, Barthel *et al.* [7] also argue that the hardness of their hard 3-SAT planted ensemble is predicated on the existence of a first-order ferromagnetic transition. We confirm the results of the TAP analysis with two alternative approaches: the replica method [19] and the annealed approximation.

First-order phase transitions are well-known to exist in the  $q$ -state Potts model for  $q \geq 3$  and in some Ising-like systems such as the Blume-Capel model [20, 21] in which variables nonetheless assume more than two states. They are, however, quite unusual in the pairwise Ising model in zero field (though see [22]); in particular neither the Sherrington-Kirkpatrick [23] nor the Edwards-Anderson [19] spin glasses exhibit such a transition. The Hopfield model [24] displays a first-order transition between the spin glass and retrieval phases when a relatively small number of patterns ( $\alpha < 0.05$ ) are stored [25], though as discussed in Section III, the Hopfield model is less appropriate as a class of problems with controllable hardness than the Wishart planted ensemble. The puzzling presence of such a transition in our system is accounted for by noting that the couplers are correlated to enforce the existence of the planted solution rather than independently disordered. Simulating systems with a first-order transition is widely known to be challenging.

The Wishart planted ensemble is of particular interest because it shares several features with models that have been shown by Nishimori and Takada [26] to be promising candidates for exhibiting (limited) exponential speedup when simulated using so-called nonstoquastic quantum driver Hamiltonians; such systems cannot be simulated classically and hence represent a “strong” type of quantum effect. The advantage of the Wishart planted ensemble over the  $p$ -spin models considered in Ref. [26] is once again that the interactions are naturally pairwise rather than requiring terms of order  $p \geq 3$ .

While nonstoquastic devices are not available as of the writing of this paper, the presented model can also be used to explore interesting problems on near-term stoquastic devices. In particular, the combination of a rough multi-modal landscape (replica symmetry breaking) in the space orthogonal to the planted solution

alongside tunable control of the energy of the planted solution (at leading order in  $N$ ) opens up the possibility to explore the hard *population transfer* problem recently proposed by Smelyanskiy *et al.* [27]. This model is useful for distinguishing physical quantum dynamics from classical dynamics such as quantum Monte Carlo in transverse field Ising models owing to the multi-path tunneling phenomena (mini-band resonance). In principle one can prepare a state in a planted mode, tune the mode energy to equality (resonance) with other modes, and explore the rate of escape.

While this paper is mostly concerned with classical properties, these intriguing connections to both types of quantum devices will certainly be explored in future work.

Computationally, the Wishart planted ensemble emerges from our procedure for generating a certain type of random integer linear program [28] and has numerous connections with well-studied [29–33] optimization problems such as the number partitioning and subset sum [34] problems. In common to these problems, the allowable precision over the problem parameters has important influence over combinatorial properties. There is also, however a crucial distinction from these problems; when the parameter  $\alpha$  is fixed, the number of equations in the integer program scales linearly with the number of variables rather than remain constant (at unity, in the case of the previous two problems).

Random problem ensembles without a planted solution typically display a parametrized “easy-hard” difficulty transition (e.g., Refs. [35] and [30]) in their optimization variants. The Wishart ensemble, on the other hand, shows an “easy-hard-easy” character. One of the easy regimes is due to the presence of a very large number of acceptable solutions coexisting with the planted ground state, while the other is due to the planting procedure effectively constraining the search space, providing “hints” to the algorithm towards the solution. The hard regime, however, is seen to be extremely difficult: numerical experiments using a distributed, state-of-the-art parallel tempering implementation show a dramatic hardness peak for small ( $N = 32$ ) system sizes. When  $N = 64$ , parallel tempering Monte Carlo fails to locate even an approximate solution under lax and permissive target criteria within the allotted simulation time of around 11 hours on contemporary high-speed hardware.

While we derive the Wishart planted ensemble in terms of a somewhat abstract random integer program, the resultant model turns out to have a remarkable structural similarity to the Hopfield model [24, 25] of biological

neurons. There important differences, however, which lead to divergent properties between the two ensembles. In particular, rather than constructing a Hamiltonian to energetically prefer spin alignment with a set of target states called patterns, the Wishart ensemble turns out to penalize alignment with a set of suitably-chosen random directions, all of which are orthogonal to the planted solution. Intriguingly, a so-called anti-Hopfield model [36] with such reversed interactions has been devised as a model of “unlearning” paramagnetic states and thereby enhance learning for biological networks. Due to planting, the correlation structure of the Wishart ensemble interactions is different from that of the anti-Hopfield model, and consequently so is its phase behavior. In particular, the anti-Hopfield model has no first-order transition comparable to the transition into the planted solution, and the large- $\alpha$  regime is Sherrington-Kirkpatrick-like rather than ferromagnetic. Nonetheless, it is exciting that such completely different starting points as integer programming and unlearning in neural networks result in models with close connections.

The rest of the paper is structured as follows. In Sec. II we describe our procedure to generate Wishart planted instances based on random integer programming problems. We also discuss the ensemble’s computational properties and how to represent its members, whose parameters are defined to take continuous values, with finite precision. Section III analyzes the physical properties of the class; the TAP free energy (derived in the Appendix) is analyzed and shown to have global minima along a one-dimensional subspace of the set of spin magnetizations. Furthermore, it has a locally stable paramagnetic state for all temperatures when  $\alpha < 1$  and for  $T > \alpha - 1$  when  $\alpha \geq 1$ . These properties give rise to the first-order transition between the paramagnetic and planted states; we determine transition temperature in terms of  $\alpha$ . We show that as  $\alpha$  grows, the system begins to increasingly behave like a ferromagnet, i.e., with a second-order transition. The predicted first-order transition temperature is validated with extensive Monte Carlo simulation. In Sec. IV, we turn our attention to empirical algorithmic properties; under finite-precision representation, the ensemble displays an easy-hard-easy relation with respect to parallel tempering time to solution as  $\alpha$  is varied. After showing that the ensemble-averaged energies of the Wishart planted ensemble follow a gamma distribution and introducing the notion of an intrinsic search space, we analytically predict the location of the hardness peak for any target energy threshold and confirm the prediction using optimized parallel tempering

simulations. The prediction is shown to be precisely accurate even for approximate solution criteria. We show that generating difficult problems under constant precision restriction requires scaling the number of constraints in the integer program approximately linearly. The Appendices VII contain most of our calculations, and confirm the TAP results through replica analysis and an annealed approximation.

## II. THE WISHART PLANTED ENSEMBLE

### A. Generation procedure

Our goal is to construct an ensemble of zero-field Ising Hamiltonians over the  $N$ -spin complete graph with planted ground state  $\mathbf{t}$ , in other words, having the form

$$H(\mathbf{s}) = -\frac{1}{2} \sum_{i \neq j} J_{ij} s_i s_j,$$

where  $\mathbf{s}$  and  $\mathbf{t}$  refer to configurations on the  $N$ -spin Ising model configuration space  $\mathbb{S}^N \triangleq \{\pm 1\}^N$  and such that

$$H(\pm \mathbf{t}) = \min_{\mathbf{s} \in \mathbb{S}^N} H(\mathbf{s}).$$

When not explicitly stated  $\mathbf{t}$  will be taken to be the ferromagnetic ground state  $\mathbf{t} = (+1, +1, \dots, +1)$  and its  $\mathbb{Z}_2$  image, as the minimizer to such a problem can be subsequently concealed by gauge randomization.

Consider the  $N \times M$  real-valued matrix  $\mathbf{W} \in \mathbb{R}^{N \times M}$ , whose  $M \geq 1$  columns are denoted by  $\mathbf{w}^\mu$  for  $\mu = 1, \dots, M$ . The value of  $M$  turns out to modulate the ensemble properties such as thermodynamics and hardness; in this work we are primarily concerned with the regime in which  $M$  scales linearly as a constant factor of  $N$ , i.e.,  $M = \alpha N$  for  $\alpha > 0$ .

Given a desired ground state  $\mathbf{t}$ , our procedure seeks to construct a consistent homogeneous Ising-constrained linear system with  $\mathbf{s} = \pm \mathbf{t}$  as a solution, in other words, to obtain  $\mathbf{W}$  such that

$$\mathbf{W}^T \mathbf{t} = \mathbf{0}. \quad (1)$$

This is because the positive semidefinite quadratic form

$$G(\mathbf{s}) = \frac{1}{2} \mathbf{s}^T \mathbf{W} \mathbf{W}^T \mathbf{s} = \frac{1}{2} \|\mathbf{W}^T \mathbf{s}\|_2^2 \quad (2)$$

would then attain its minimum value of zero at  $\mathbf{s} = \mathbf{t}$ , and hence if we define

$$\tilde{\mathbf{J}} = -\frac{1}{N}\mathbf{W}\mathbf{W}^T = -\frac{1}{N}\sum_{\mu=1}^M \mathbf{w}^\mu \otimes \mathbf{w}^\mu \quad (3)$$

and zero its diagonal to form

$$\mathbf{J} = \tilde{\mathbf{J}} - \text{diag}(\tilde{\mathbf{J}}), \quad (4)$$

the Hamiltonian

$$H(\mathbf{s}) = -\frac{1}{2}\mathbf{s}^T \mathbf{J} \mathbf{s} \quad (5)$$

attains its ground state at  $\mathbf{s} = \mathbf{t}$  with energy

$$\begin{aligned} H(\mathbf{t}) &= -\frac{1}{2}\mathbf{t}^T \mathbf{J} \mathbf{t} \\ &= \frac{1}{2}\text{Tr}(\tilde{\mathbf{J}}), \end{aligned}$$

where the property that  $s_i^2 = 1$  for  $\mathbf{s} \in \mathbb{S}^N$  has been used. The scaling by  $1/N$  in the definition of  $\tilde{\mathbf{J}}$  is to make the energy extensive, i.e., scaling linearly with system size.

We obtain the linear system by individually generating the  $M$  columns  $\{\mathbf{w}^\mu\}, i \in 1, \dots, M$  of  $\mathbf{W}$  such that  $\langle \mathbf{w}^\mu, \mathbf{t} \rangle = 0$ . We propose a simple projective method for efficiently generating correlated Gaussian variates satisfying the summation and other desirable properties. More precisely, the column vectors are set to be distributed as

$$\mathbf{w}^\mu \sim \mathcal{N}(\mathbf{0}, \Sigma), \quad (6)$$

where the covariance matrix is given by

$$\Sigma = \frac{N}{N-1} \left[ \mathbf{I}_N - \frac{1}{N} \mathbf{t} \mathbf{t}^T \right] \quad (7)$$

with  $\mathbf{I}_N$  the  $N$ -dimensional identity matrix. In other words, for each column vector  $\mathbf{w}$ , all elements have unit variance, and for all variable pairs  $i \neq j$ , the covariances are

$$\mathbb{E}[w_i w_j] = -\frac{t_i t_j}{N-1}.$$

Note that  $\text{rank}(\Sigma) = N-1$  as expected; given any  $N-1$  components of  $\mathbf{w}$ , the remaining one follows deterministically. To generate the column vectors, we first determine

the square root of  $\Sigma$ , i.e.,  $\Sigma^{\frac{1}{2}}$  such that  $\Sigma = \Sigma^{\frac{1}{2}} \Sigma^{\frac{1}{2}}$ , to be

$$\Sigma^{\frac{1}{2}} = \sqrt{\frac{N}{N-1}} \left[ \mathbf{I}_N - \frac{1}{N} \mathbf{t} \mathbf{t}^T \right]. \quad (8)$$

We then iterate over the loop described in Algorithm 1:

---

**Algorithm 1** Wishart Planted Ensemble Generator

---

```

for  $\mu = 1, \dots, M$  do
    Sample uncorrelated Gaussian  $\mathbf{z}^\mu \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_N)$ 
     $\mathbf{w}^\mu \leftarrow \Sigma^{\frac{1}{2}} \mathbf{z}^\mu$ 
end for

```

---

One can readily verify that  $\langle \mathbf{w}^\mu, \mathbf{t} \rangle = 0$  for all  $\mu$  and that  $\mathbf{w}^\mu$  is distributed according to Eq. (6). While the components of  $\mathbf{w}$  are correlated, the Gaussian nonetheless has strong structure that simplifies the subsequent analysis. Following an appropriate gauge transformation to the ferromagnetic state, the elements of  $\Sigma$  imply that the distribution is *exchangeable*, i.e., invariant to a permutation of the components. Exchangeability is a stronger property than stationarity, which merely requires the covariance of components  $i$  and  $j$  to depend only on  $|i - j|$ . The property is used in Sec. III A when deriving the Thouless-Anderson-Palmer (TAP) free energy and again in Sec. IV A when obtaining the ensemble energy distribution.

The random matrix  $\mathbf{W}\mathbf{W}^T$  follows a Wishart distribution, a well-studied matrix generalization of the  $\chi^2$  distribution; in light of this we call our problem class the Wishart planted ensemble (WPE). When  $M < N$ , the support of the Wishart density lies on a low-dimensional subspace of  $N \times N$  matrices [37]. As we see in Sec. III B, its spectral distribution is a key feature underlying the phase behavior. We note also that while sampling  $\mathbf{w}^\mu$  according to the Gaussian  $\mathcal{N}(\mathbf{0}, \Sigma)$  simplifies the analysis it is by no means necessary in practice for our results to hold in the large  $N$  limit. For example if  $\{\mathbf{z}^\mu\}$  used to obtain  $\{\mathbf{w}^\mu\}$  were  $N$ -vectors consisting of independent and uniform  $\{\pm 1\}$  variates rather than uncorrelated Gaussians, central limit arguments show that  $\mathbf{W}\mathbf{W}^T$  is nonetheless asymptotically Wishart.

## B. Computational properties

Before proceeding to an examination of the WPE thermodynamics, we discuss its properties from a computa-

tional perspective in light of its interpretation as a constrained homogeneous linear system. Readers familiar with related settings such as linear error correcting codes should bear in mind that arithmetic here is over the real numbers rather than a finite field such as  $\text{GF}(2)$ .

Given a matrix  $\mathbf{W}$ , the task of finding the ground state of Eq. (5) is equivalent to finding a solution to the following NP-hard problem called *integer programming feasibility*

$$\begin{aligned} &\text{solve } \mathbf{W}^T \mathbf{s} = 0 \\ &\text{subject to } \mathbf{s} \in \mathbb{S}^N. \end{aligned} \quad (9)$$

Suspending for a moment the fact that  $M$  scales with  $N$  in the WPE, we can obtain a sense for how it may impact problem difficulty. If  $\mathbf{W}$  consists of  $M < N$  independent columns, then  $\dim(\text{null}(\mathbf{W}^T)) = N - M$ . The dimension of the nullspace of  $\mathbf{W}^T$  implies the search space for potential Ising state solutions to Eq. (9), and so the larger it is, the more difficult the problem may be guessed to be. In particular, when  $\mathbf{W}^T$  consists of a single row ( $M = 1$ ) and hence an  $N - 1$  dimensional nullspace, the problem may be surmised to be maximally hard. When  $\mathbf{W}$  is specified with relatively low precision, this turns out to not be the case; exponentially many solutions other than  $\pm \mathbf{t}$  overwhelmingly appear as  $N$  increases and so locating any such satisfying  $\mathbf{s}$  can be quite easy. This is reflected in the Ising Hamiltonian (5); when  $M = 1$ ,  $\mathbf{J}$  can be verified to be fully-frustrated, which gives rise to tremendous low-energy degeneracy.

At the other extreme in which  $M$  assumes large values, the nullspace of  $\mathbf{W}^T$  becomes one-dimensional and hence the two Ising solutions to Eq. (9) are trivially recovered by inspection from a vector spanning the nullspace. In the Hamiltonian in Eq. (5), making  $M$  large results in a ferromagnetic system. To see this, we note that

$$\tilde{\mathbf{J}} = -\frac{1}{N} \mathbf{W} \mathbf{W}^T \quad (10)$$

$$= -\frac{\alpha}{M} \mathbf{W} \mathbf{W}^T \quad (11)$$

$$\rightarrow -\alpha \mathbf{\Sigma}, \quad (12)$$

where the limit follows from the law of large numbers. From the covariance matrix  $\mathbf{\Sigma}$ , the couplers thus uniformly approach

$$J_{ij} = \alpha \frac{t_i t_j}{N - 1},$$

implying that the system reduces to a gauge-transformed and rescaled Curie-Weiss ferromagnet, whose ground

state is easy to find due to the lack of frustration. Hence one may reasonably guess that for a given allowable precision, the most difficult problems occur for some intermediate value of  $M$ . This easy-hard-easy profile is shown to indeed hold and is discussed in Sec. IV, in which we make more explicit the role of  $\text{null}(\mathbf{W}^T)$  and develop the conjecture that the difficulty peak occurs at the value of  $M$  in which the fewest number of solutions occur relative to the volume of the nullspace.

A prototypical special case of Eq. (9), corresponding to  $M = 1$ , is the subset sum problem; its decision variant, that of establishing whether a satisfying  $\mathbf{s}$  to Eq. (9) exists, was one of Karp's [38] original 21 NP-complete problems. We note that  $\mathbf{W}$  must be specifiable to arbitrary precision for the optimization variant in Eq. (9) to avoid efficient solution via the technique of dynamic programming as the complexity of this algorithm scales exponentially in the number of bits needed to specify the coefficients. This requirement holds more generally: when the value of  $M$  is fixed to any integer and the matrix elements belong to a finite set, integer programming problems of the form of Eq. (9) can be solved in pseudopolynomial time using the same technique [39].

Subset sum problems with feasible solutions have been of particular interest due to their complexity underpinning the security of an early family of cryptographic systems [40]. Remarkably, a method has been devised [33] for solving with high probability a family of low-density random subset sum problems, in which the maximum element of the single-column  $\mathbf{W}$  is large compared to  $N$ , based on an idea known as lattice basis reduction [41].

A well-studied further specialization of the subset sum problem is known as the number partitioning problem, corresponding to the task of partitioning a base set of *positive* integers into two blocks with sums of minimum absolute difference (or “discrepancy”). When the integers are chosen independently from a uniform distribution (the problem does not respect a planted solution), number partitioning displays several interesting properties, in particular an algorithmic easy-hard phase transition [30, 31, 35, 42]. More specifically, if the positive integers forming the base set are bounded by  $2^{\kappa N}$  for a fixed  $\kappa > 0$ , then when  $\kappa > \kappa_c$ , partitions with a discrepancy of zero exist with vanishing probability, and typical instances become difficult for known heuristic and complete algorithms. Conversely, when  $\kappa$  lies below this threshold, partitions with zero discrepancy are abundant, and the problems are easily solved.

While the subset sum and number partitioning problems

bear obvious connections to the model presented in this work, it is apparent that our assumptions and focus are different. Broadly speaking, in the WPE a state must be found which now simultaneously satisfies the  $M$  relations in the integer program; the fact that  $M$  scales as  $\alpha N$  rather than remaining fixed introduces consequences in the algorithmic hardness properties.

### C. Representing a WPE instance

The WPE construction presented in Sec. II A was defined in terms of Gaussian variables; in a practical implementation, however, one must contend with finite precision and generally cannot represent the required continuous-valued parameters. In the WPE, this can be dealt with in one of two ways. The first, which will be examined in this section, is to maintain the correlation structure defined by  $\Sigma$  but replace the Gaussian variates  $\mathbf{z}$  with a discrete zero-mean, unit variance ensemble; the simplest such choice is to have  $\mathbf{z}$  uniformly and independently take the two values  $\{\pm 1\}$  (sometimes called a Rademacher distribution). This turns out to allow an *exact* representation of the problem parameters using a logarithmic (in  $N$ ) number of bits; further, as mentioned in Sec. II A, the coupler matrix is nonetheless asymptotically Wishart and the same physical properties derived in this paper result.

The second and more heuristic way to represent a problem is to simply round the parameters to the closest machine-representable number. This introduces numerical errors; for example the planted solution may no longer have its theoretically-intended energy. Consequently, one must introduce a tolerance on what defines a “solution.” We use this approach in our algorithmic hardness experiments, discussed in Sec. IV, as it allows us to leverage our existing high-performance parallel tempering codes. Nonetheless, using the theoretically-derived energy histogram, we are able to account for the observed hardness peak under this approximate representation.

We now discuss the exact discretized WPE over a Rademacher distribution, i.e., where  $\{z_i^\mu\}$  are independently and uniformly in  $\{-1, 1\}$  rather than drawn from  $\mathcal{N}(0, 1)$  for  $i \in \{1, \dots, N\}$  and  $\mu \in \{1, \dots, M\}$ . We first rewrite

$$\Sigma^{\frac{1}{2}} = \frac{1}{\sqrt{N(N-1)}} \mathbf{A},$$

where the integer-valued matrix

$$\mathbf{A} \triangleq N \left[ I_N - \frac{1}{N} \mathbf{t} \mathbf{t}^T \right],$$

i.e.,

$$A_{ij} = \begin{cases} N-1 & i=j \\ -1 & i \neq j \end{cases}.$$

Because

$$\mathbf{w} = \frac{1}{\sqrt{N(N-1)}} \mathbf{A} \mathbf{z},$$

then up to the leading constant, the elements of  $\mathbf{w}$  may assume values in the set of  $2N-1$  equally-spaced integers

$$\mathcal{S}_{\mathbf{w}} \triangleq \{-2(N-1), \dots, -2, 0, 2, \dots, 2(N-1)\}.$$

Thus, in the integer programming formulation, the Rademacher-discretized WPE takes approximately  $\log(2N-1)$  bits to encode the parameters.

Obtaining the required precision in the Hamiltonian formulation (i.e., on  $\mathbf{J}$ ) is more messy but analogous. Recalling that

$$J_{ij} = -\frac{1}{N} \sum_{\mu=1}^M w_i^\mu m_j^\mu,$$

then using the previous restriction on the values of  $\mathbf{w}$  we can show that

$$N^2(N-1)J_{ij} \in \mathcal{S}_{\mathbf{J}},$$

where

$$\mathcal{S}_{\mathbf{J}} \triangleq \{-4M(N-1)^2, \dots, -4, 0, 4, \dots, 4M(N-1)^2\}.$$

Not all elements in this set of integers are actually attainable by  $J_{ij}$ , but it provides a useful upper bound on the number of possible values and shows that representing  $J_{ij}$  in the Rademacher-discretized WPE takes no more than on the order of  $\log(M) + 2 \log(N)$  bits, which when  $M = \alpha N$  is  $O(\log N)$ .

### III. THERMODYNAMIC PROPERTIES

Analyzing long-range disordered systems has a rich history in statistical mechanics [43]. The Sherrington-Kirkpatrick (SK) model [23], a fully-connected Ising

model with independently-sampled Gaussian bond strengths, is a prototypical example of such systems. The replica method [19] is a powerful framework for performing such analyses, and has yielded great successes such as the currently accepted correct solution to the SK model [44]. This approach is pursued in Appendix VII E where we are able to recover the transition properties developed in this section, and through connections with the anti-Hopfield model identify some additional interesting transitions within the model.

A different and in many senses complementary approach to the analysis weakly-coupled, fully-connected disordered systems is due to Thouless, Anderson, and Palmer [15]. The TAP equations are a set of nonlinear relations satisfied by the local magnetizations for a given instance. They can be arrived at in one of several ways [45]. For the SK model, they are often interpreted as correcting the “naïve” mean field equations with a so-called Onsager reaction term. The approach of Plefka [46] arrives at the TAP equations by second-order expansion of the free energy at constant magnetization, which turns out to have an appealing information geometric interpretation [47]. In this section we use another approach called the cavity method [48].

Determination of the TAP equations for the WPE is somewhat complicated by the fact that  $J_{ij}$  are not independent variates as they are for the SK model. The TAP equations for systems with correlated  $\mathbf{J}$  have been determined in the past: a notable example, which turns out to have a close connection to our ensemble, is the Hopfield model [24] of a biological neural network. The couplings of the Hopfield model are given by

$$J_{ij} = \frac{1}{N} \sum_{\mu=1}^p \xi_i^\mu \xi_j^\mu,$$

where the  $p$  vectors  $\{\xi^\mu\}$ , known as patterns to be stored for later retrieval, consist of  $N$  *independent* zero-mean binary random variables. The physics of the Hopfield model was first studied with the replica approach by Amit *et al.* [25, 49]; the TAP equations were derived by Mézard *et al.* [48] using the cavity method, but yielded results inconsistent with the replica analysis. TAP equations consistent with Ref. [49] were obtained by Nakanishi and Takayama [50] using Plefka’s method, and subsequently by Shamir and Sompolinski [51] via an elegant cavity approach.

The connection of the WPE to the Hopfield model is apparent if we express the elements of  $\tilde{\mathbf{J}}$  defined in Eq. (3)

as

$$\tilde{J}_{ij} = -\frac{1}{N} \sum_{\mu=1}^M w_i^\mu w_j^\mu.$$

An obvious difference from the Hopfield model is the presence of the leading negation. Consequently, while the Hopfield Hamiltonian tends to favor spin configurations aligned with the patterns  $\{\xi^\mu\}$ , the WPE Hamiltonian penalizes configurations overlapping with the directions  $\{\mathbf{w}^\mu\}$ . An additional distinction, however, is that while the  $M$  vectors are independently drawn from the previously-defined Gaussian distribution, the components of each vector are now correlated, a property which emerged due to the solution planting procedure.

Remarkably, a model defined by negating the sign of the Hopfield model Hamiltonian (only the first of the two differences above) has been proposed as a model of “unlearning” paramagnetic configurations and thereby enhancing learning for biological networks. A replica analysis of this “anti-Hopfield” model was undertaken by Nokura [36]. While the additional “layer” of correlations and the presence of a planted solution in the WPE turns out to lead to very different behavior from that of the anti-Hopfield model, it is nonetheless quite interesting that the two models arrived at by very distinct paths, namely an Ising-based encoding of an integer linear program with known solution in our case and a model hoping to improve learning in neural networks in the anti-Hopfield case, turn out to have a considerable resemblance.

Another Ising ensemble related to the WPE is the random orthogonal model (ROM) proposed by Parisi and Potters [52], where the  $\mathbf{J}$  matrices are generated by uniformly sampling an orthogonal matrix  $\mathbf{O}$ , forming diagonal matrix  $\mathbf{D}$  whose elements  $D_{ii} \in \{\pm 1\}$  (often additionally assumed to have a trace of zero), and setting  $\mathbf{J} = \mathbf{O} \mathbf{D} \mathbf{O}^T$ . Having been devised with different aims, this model is also quite different from the WPE. First, no planting takes place, so the ground state is uncontrolled. Further, the ground state energy is only known if an Ising-feasible state happens to be an eigenvector of  $\mathbf{J}$ , which become exceedingly unlikely for even moderately-sized systems. Finally, the eigenvalue distributions of the  $\mathbf{J}$  matrices not the same, nor are the TAP equations presented in Ref. [52]. The systems thus have quite differing thermodynamic properties; in particular, a first-order transition analogous to the planting transition is absent in the ROM.

It is natural to wonder about the suitability of the orig-



inal Hopfield model to the task at hand. Unfortunately, it is not possible to generate sufficiently rugged energy landscapes while maintaining control over the ground states using this model. Hence, while the Hopfield model is an appealing abstraction of an associative memory, it is not as well-suited to usage as a tunable planted optimization objective. The Hopfield model does not exhibit, for any setting, a persistent metastable paramagnetic state analogous to that of the  $\alpha < 1$  WPE. In contrast to the WPE in which the planted ground state is undisturbed while a rough energy landscape is induced in the orthogonal subspace, ruggedness in the Hopfield model arises due to interference among the stored patterns, causing undesired spurious minima [25]. For values of  $\alpha$  less than around 0.05, a first-order transition between a spin glass and "retrieval" phase takes place, but finding one of the embedded patterns is quite easy in this case (as it should be for the model's intended purpose). At somewhat larger values of  $\alpha$ , the model continues to function as an associative memory, though the patterns are only assured of being local optima rather than ground states. When  $\alpha > 0.138$ , only a second-order paramagnetic to spin glass transition takes place, and control over even the local minima is completely lost.

### A. TAP equations

We derive the TAP equations for the WPE following the two-step cavity approach of Shamir and Sompolinski [51], making appropriate adaptations to deal with the correlations among the components of each  $\mathbf{w}$ . The complete calculation is shown in Sec. VII A. Alternatives to the TAP method based on the annealed approximation and replica method are considered in Appendices VIID and VIIE, yielding consistent results. In these appendices the connection with the anti-Hopfield model is explored, as is the notion of a planted solution with tuned energy that might permit other uses, such as the population transfer experiment discussed in the introduction.

Let  $m_i = \langle s_i \rangle$  and  $\mathbf{m}$  be the vector of all magnetizations. Define

$$q = \frac{1}{N} \sum_{i=1}^N m_i^2$$

and

$$V = \frac{\alpha(1-q)}{1+\beta(1-q)}.$$

The TAP equations for the WPE are

$$m_i = \tanh \left[ \beta \left( \sum_{j \neq i} J_{ij} m_j - \beta V m_i \right) \right]. \quad (13)$$

Solutions to Eq. (13) are stationary points of the following TAP free energy:

$$F_{\text{TAP}}(\mathbf{m}) = -\frac{1}{2} \sum_{i \neq j} J_{ij} m_i m_j - \frac{1}{\beta} \sum_{i=1}^N S(m_i) - \frac{1}{2} \alpha N \left[ (1-q) - \frac{1}{\beta} \log \left( 1 + \beta(1-q) \right) \right], \quad (14)$$

where the local entropy terms are

$$S(m_i) = -\frac{1+m_i}{2} \log \frac{1+m_i}{2} - \frac{1-m_i}{2} \log \frac{1-m_i}{2}.$$

In general, determining the global minima of  $F_{\text{TAP}}$  is a difficult task as  $\mathbf{m}$  lies in an  $N$ -dimensional space subject to the bound constraints

$$\mathbf{m} \in [-1, 1]^N.$$

In our case however, the existence of the planted solution considerably simplifies things by assuring that the free-energy minimum necessarily occurs along the ground state direction:

$$\min_{\mathbf{m} \in [-1, 1]^N} F_{\text{TAP}}(\mathbf{m}) = \min_{m \in [-1, 1]} F_{\text{TAP}}(m\mathbf{t}).$$

To see this, consider the restriction of  $\mathbf{m}$  to the spherical shell given by

$$\left\{ \mathbf{m} \mid \frac{1}{N} \sum_{i=1}^N m_i^2 = q_0 \right\}$$

for some  $q_0 \in [0, 1]$ , or alternatively,

$$\mathbf{m} = \mathbf{e} \sqrt{N q_0}$$

for arbitrary unit vector  $\mathbf{e}$ . The claim is that

$$\mathbf{e} = \pm \frac{\mathbf{t}}{\sqrt{N}}$$

minimizes  $F_{\text{TAP}}$  on the shell; from Eq. (14), the term

$$-\frac{1}{2} \sum_{i \neq j} J_{ij} m_i m_j$$

is minimized for  $\mathbf{m} = \pm \sqrt{q_0} \mathbf{t}$  due to the planting procedure. Further, the term

$$-\sum_{i=1}^N S(m_i)$$

is minimized on the shell at the  $2^N$  points at which the  $\{m_i\}$  have the same magnitude, which holds when  $\mathbf{m} = \pm \sqrt{q_0} \mathbf{t}$ , and the final term in  $F_{\text{TAP}}$  is independent of  $\mathbf{m}$  on the sphere. The global optimum, which is the minimum over all the shells' minimizers, thus occurs along

$$\mathbf{m} = m \mathbf{t} = \pm \sqrt{q} \mathbf{t}.$$

Recall that

$$H(\mathbf{t}) = -\frac{1}{2} \sum_{ij} J_{ij} t_i t_j = \frac{1}{2} \text{Tr}(\tilde{J})$$

and that

$$\tilde{J}_{ii} = -\frac{1}{N} \sum_{\mu=1}^M (w_i^\mu)^2.$$

Using the fact that  $S(mt_i) = S(m)$  for  $t_i \in \{\pm 1\}$  we define the one-dimensional TAP free energy as  $F(m) \triangleq F_{\text{TAP}}(m\mathbf{t})$ , i.e.,

$$F(m) \triangleq -\frac{1}{2N} m^2 \sum_{\mu=1}^M \sum_{i=1}^N (w_{ii}^\mu)^2 - \frac{N}{\beta} S(m) - \frac{1}{2} \alpha N (1 - m^2) + \frac{\alpha N}{2\beta} \log(1 + \beta(1 - m^2)). \quad (15)$$

Consider the difference

$$\frac{m^2}{2} \left[ \alpha N - \frac{1}{N} \sum_{\mu=1}^M \sum_{i=1}^N (w_{ii}^\mu)^2 \right]$$

present in Eq. (15). Using the higher-order moments of correlated Gaussians, one can show that despite the dependence among the components of  $\mathbf{w}^\mu$ , the term scaling

$\frac{m^2}{2}$  is at most  $O(\sqrt{M})$ . Normalizing by  $N$  to obtain the free energy per spin, we obtain

$$f(m) = -\frac{m^2}{2N} O(\sqrt{\alpha N}) - \frac{1}{2} \alpha - \frac{1}{\beta} S(m) + \frac{\alpha}{2\beta} \log(1 + \beta(1 - m^2)). \quad (16)$$

In the thermodynamic limit, the first term vanishes; neglecting the constant term of  $-\frac{1}{2}\alpha$ , the global free energy minima of Eq. (14) can thus be found as those of the function

$$\tilde{f}(m) \triangleq -\frac{1}{\beta} S(m) + \frac{\alpha}{2\beta} \log(1 + \beta(1 - m^2)). \quad (17)$$

The minima of  $\tilde{f}$  must be stationary points of Eq. (17), which are attained for solutions of the equation

$$m = \tanh \frac{\alpha \beta m}{1 + \beta(1 - m^2)}. \quad (18)$$

## B. Thermodynamic properties of the WPE

The WPE displays phase properties quite surprising in the Ising model. Rather than undergo an SK-like spin-glass transition typical of many disordered systems, the presence of a planted solution gives rise to a thermal first-order phase transition. This transition in the WPE occurs at an  $\alpha$ -dependent temperature  $T_c$  such that two stable and equally-contributing states exist in the free energy. As expected, at high temperature the disordered paramagnetic state ( $\mathbf{m} = \mathbf{0}$ ) is the unique global minimizer to  $F_{\text{TAP}}$ . At a critical temperature  $T_c$ , this state remains globally minimizing but is no longer unique; another state  $\mathbf{m} = m\mathbf{t}$  with  $m \neq 0$  is also optimal. Below  $T_c$ , the paramagnet ceases to be the minimum, but whether  $\mathbf{m} = \mathbf{0}$  remains a local minimum (metastable state) to  $F_{\text{TAP}}$  or an unstable state (saddle point or maximizer) depends on  $\alpha$  and the temperature. We find the fascinating result that when  $\alpha < 1$ , the paramagnetic state is locally stable at all nonzero temperatures. As will be discussed, this property has considerable algorithmic implications.

The transition is first-order because of the discontinuity in the minimizing  $\mathbf{m}$  and in the derivative of the free-energy minimum with respect to  $T$ ; consequently the internal energy also undergoes a jump at  $T_c$  as we will see. Remarkably, for any finite  $\alpha$ , the transition is always technically first-order, in the sense that there is

a discontinuity in the log partition function derivative. The magnitude of this discontinuity decreases with  $\alpha$ , and the transition gradually segues from first to second order.

### 1. Stability of $\mathbf{m} = \mathbf{0}$

The stability of  $\mathbf{m} = \mathbf{0}$  can be ascertained from the Hessian of  $F_{\text{TAP}}$ , i.e.,

$$\mathbf{H}(\mathbf{m}) \triangleq \frac{\partial^2 F_{\text{TAP}}}{\partial \mathbf{m}^2}.$$

The point  $\mathbf{m} = \mathbf{0}$  is a local minimum of  $F_{\text{TAP}}$  if and only if  $\mathbf{H}(\mathbf{0})$  is positive definite. It turns out that the spectral distribution of the random matrix  $\mathbf{J}$  is key in determining the relation between the definiteness of  $\mathbf{H}(\mathbf{0})$  and  $T$  at given  $\alpha$ . The limiting eigenvalue distribution of  $\mathbf{J}$  is calculated in Sec. VII B; somewhat surprisingly, it is closely related to the Marchenko-Pastur [53] law for the spectrum of Wishart matrices constructed with *uncorrelated* Gaussians. The important aspect for now is that  $-\alpha$  is always (asymptotically) the smallest eigenvalue of  $-\mathbf{J}$ .

Computing the partial derivatives from the definition of  $F_{\text{TAP}}$  in Eq. (14) we obtain the Hessian matrix elements

$$H_{ij} = \begin{cases} \frac{1}{\beta} \text{sech}^2(m_i) - \frac{2\alpha\beta m_i^2}{N(1+\beta(1-q))^2} + \frac{\alpha\beta(1-q)}{1+\beta(1-q)} & i = j \\ -J_{ij} - \frac{2\alpha\beta m_i m_j}{N(1+\beta(1-q))^2} & i \neq j \end{cases}.$$

At  $\mathbf{m} = \mathbf{0}$  we thus have

$$H_{ij} = \begin{cases} \frac{1}{\beta} + \frac{\alpha\beta}{1+\beta} & i = j \\ -J_{ij} & i \neq j \end{cases}$$

or compactly, the Hessian at  $\mathbf{m} = \mathbf{0}$

$$\mathbf{H}(\mathbf{0}) = -\mathbf{J} + c(\alpha, \beta)\mathbf{I} \quad (19)$$

with

$$c(\alpha, \beta) = \frac{1 + \beta(1 + \alpha)}{\beta(1 + \beta)}.$$

$\mathbf{H}(\mathbf{0})$  is positive definite when  $c(\alpha, \beta)$  is large enough to shift the eigenvalues of  $-\mathbf{J}$  to all be positive. From the spectral property mentioned above, stability hence occurs when

$$c(\alpha, \beta) \geq \alpha.$$

For a given  $\alpha$ , this condition is equivalent to

$$(T + 1)(\alpha - T) \leq \alpha$$

or

$$T(T + 1 - \alpha) \geq 0.$$

Obviously,  $T$  is nonnegative; when  $\alpha < 1$  the relation thus holds for all  $T$  while when  $\alpha \geq 1$ , it is satisfied when  $T \geq \alpha - 1$ .

To summarize, when  $\alpha < 1$ , the point  $\mathbf{m} = \mathbf{0}$  is a locally stable stationary point to the TAP free energy at any temperature; on the other hand when  $\alpha \geq 1$ , the Hessian becomes indefinite at temperature  $T_u = \alpha - 1$ . For temperatures below  $T_u$ , note that the eigenvalue of  $\mathbf{H}$  corresponding to eigenvector  $\mathbf{t}$  (the planted solution) becomes negative, which implies that the paramagnetic solution becomes unstable along this direction. Interestingly, the curvature of  $F_{\text{TAP}}$  along  $\mathbf{t}$  is of largest magnitude for  $\alpha \geq 1$  at the temperature such that  $c(\alpha, \beta)$  is minimal. In terms of  $T$ ,

$$c(\alpha, T) = T + \frac{\alpha}{T + 1}$$

which is smallest when  $T = -1 + \sqrt{\alpha}$ . As local search heuristics follow free-energy gradients in their quest for the solution, it seems reasonable that this temperature is in some sense optimal to start with when searching for the ground state in the  $\alpha \geq 1$  (easy) regime. A further special interpretation of this temperature relates to the anti-Hopfield model, as discussed in Appendix VII E.

### 2. First-order phase transition

The first-order transition for given  $\alpha$  occurs at a  $T_c$  such that  $\tilde{f}(0) = \tilde{f}(m)$  for  $m \neq 0$ . For any  $T$ , the stationary points of  $\tilde{f}$  are determined numerically by iteratively solving the saddle point equation [Eq. (18)]; using binary search in  $T$ , we can then localize the transition temperatures at any  $\alpha$ . Figure 1 shows the relation of  $T_c$  with  $\alpha$ , and also shows for  $\alpha \geq 1$  the paramagnetic instability point  $T_u = \alpha - 1$ . It is clear that  $T_c > T_u$  uniformly but as  $\alpha$  increases the two temperatures converge; this in turn constrains the jump magnitude between the two sides of the transition. While  $T_c$  has no closed-form expression in  $\alpha$ , we can obtain a lower bound  $\hat{T}_c$  as

$$T_c \geq \hat{T}_c \triangleq 2^{2/\alpha} - 1 \quad (20)$$

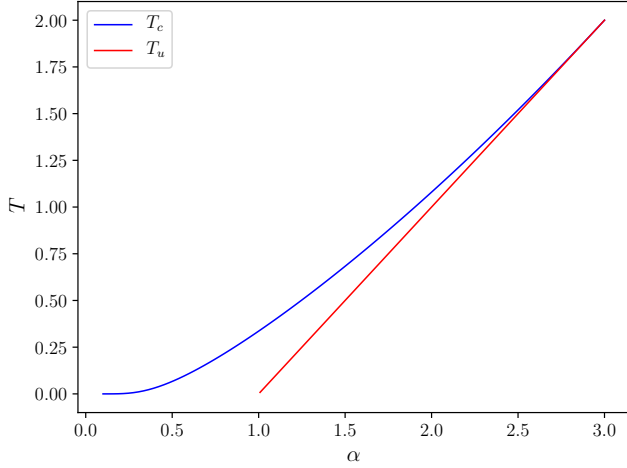


FIG. 1. Transition temperatures for the Wishart planted ensemble as a function of  $\alpha$ .  $T_c$  (solid blue) refers to the first-order (coexistence) phase transition temperature, while for  $\alpha \geq 1$ ,  $T_u$  (solid red) marks the instability of the paramagnetic phase and a ferromagnetic transition. Note that as  $\alpha$  increases past 1,  $T_u$  and  $T_c$  converge, which is expected from the construction procedure. The values of  $T_c$  in the low- $\alpha$  portion of the function are small but nonzero; they are well-approximated there by  $T_c \approx 2^{2/\alpha} - 1$ .

which serves as a surprisingly accurate approximation for small  $\alpha$  but deteriorates as  $\alpha$  gets larger than 1.

We illustrate these results by plotting  $\tilde{f}(m)$  at representative temperatures for a few values of  $\alpha$ . Consider first the relatively small value of  $\alpha = 0.25$ , with landscape illustrated in Fig. 2. The first-order transition at  $T_c = 0.004$  is clearly visible in the middle panel as  $m = 0$  is now a coexisting global minimum with the other minima close to the endpoints. Remarkably, when  $T \approx 0.003$ , shown in the bottom panel,  $m = 0$  remains locally optimal.

In Fig. 3 we see similar behavior with an increased  $T_c \approx 0.187$  and a persistent metastable  $m = 0$  state at low temperature for  $\alpha = 0.75$ .

From an algorithmic perspective, the low-temperature stability of  $m = 0$  is quite intriguing as it is widely believed to correlate with genuine combinatorial hardness, sounding the death knell for heuristic approaches attempting to locally optimize trial configurations. This category of algorithms includes workhorses such as simulated annealing and parallel tempering Monte Carlo, which exploit correlations in the energy landscape to search by performing biased random walks. Paramag-

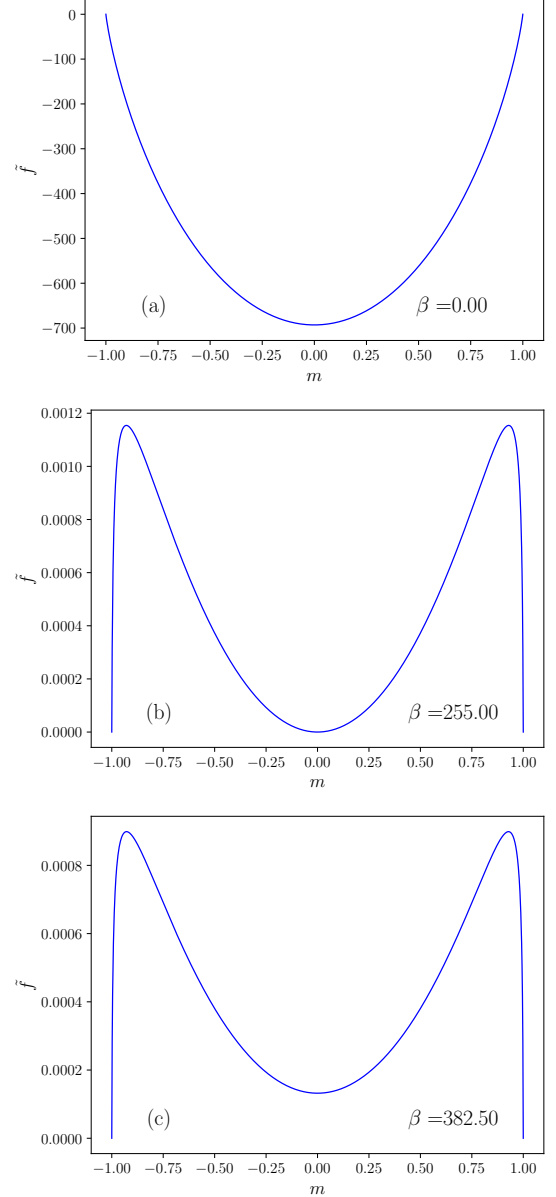


FIG. 2. Function  $\tilde{f}$ , whose global minima correspond to those of the TAP free energy, for a Wishart planted problem with  $\alpha = 0.25$ . In (a), we see the high-temperature regime, with a single minimum at  $m = 0$ . Panel (b) shows the first-order transition, occurring at  $T_c = 0.004$  and characterized by equal contribution of  $m = 0$  and the minima close to  $m = \pm 1$ , while panel (c) displays the low-temperature phase ( $T \approx 0.003$ ). Note that  $m = 0$  remains metastable in spite of the low temperature, while the free-energy minima lie around  $m = \pm 1$  corresponding to the planted solution.

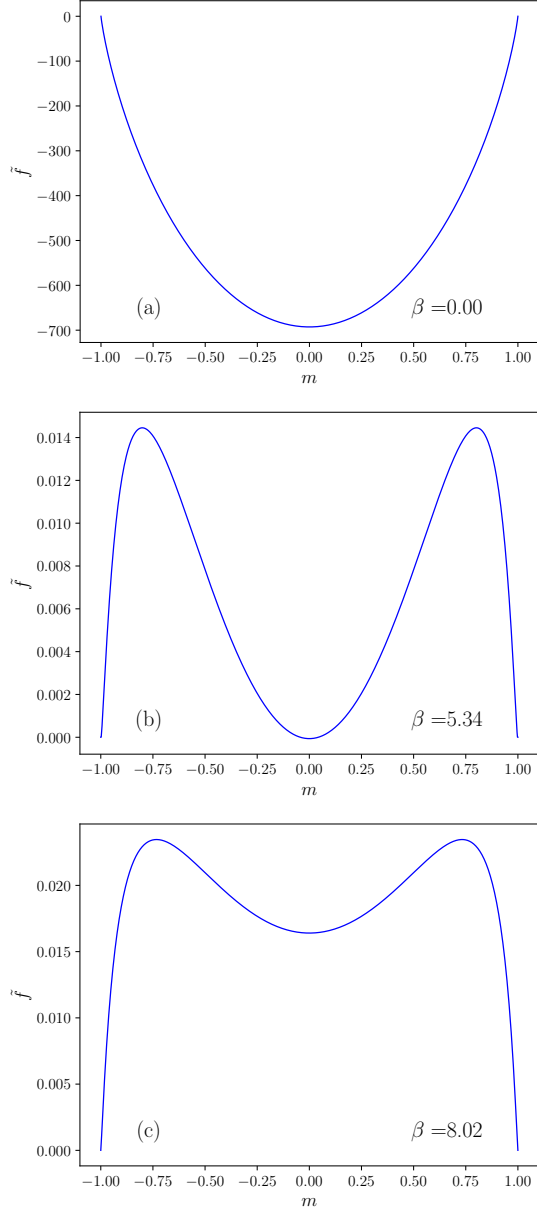


FIG. 3. Function  $\tilde{f}$  for a Wishart planted problem with  $\alpha = 0.75$ . In panel (a) we show the high-temperature regime. Panel (b) shows the first-order transition at  $T_c = 0.187$ . Panel (c) illustrates the low-temperature ( $T \approx 0.125$ ) phase, with  $m = 0$  remaining metastable as it did for  $\alpha = 0.25$  shown in Fig. 2. Note further that the sizes of the regions between the *maxima* and the boundaries are larger than they were for  $\alpha = 0.25$ , suggesting that the ground state has a wider basin of attraction, resulting in this being a computationally easier problem.

netic stability suggests that a problem is hard because it implies that cooling an initially-disordered configuration will overwhelmingly lead to states that are also disordered. There is no exploitable information within the landscape guiding the dynamics to the correct region of the state space, and the only hope is to begin the search from the appropriate basin. Disordered states form the vast bulk of the configuration space, however, so the probability of a correct initialization decreases exponentially with system size. While all problems with  $\alpha < 1$  exhibit this feature, the specific value of  $\alpha$  turns out to modulate the size of the planted solution basin, i.e., all states from which the ground state can be reached at reasonable cost, which of course exerts a critical effect on the observed performance of heuristics. This aspect can be observed by comparing the intervals between the endpoints and *maxima* of  $\tilde{f}$  in Figs. 2 and 3 and noting them to be wider for  $\alpha = 0.75$ . Broadly speaking, small values of  $\alpha$  lead to smaller basins and hence, one would surmise, more difficult problems. This assertion assumes, however, that the vectors  $\{\mathbf{w}^\mu\}$  are comprised of real numbers and specifiable to arbitrary numerical precision. When the precision is bounded, the interpretation is more involved and is discussed in detail in Sec. IV.

Figure 4 shows how  $\tilde{f}$  is affected by  $T$  for a value of  $\alpha = 1.5$ , a regime in which we expect  $m = 0$  to become unstable. The first-order transition is still readily apparent at  $T_c = 0.68$ . For temperatures lower than  $T_c$  but higher than  $T_u = 0.5$ , the paramagnet remains a local minimum. Finally, when cooled to  $T_u = 0.5$ ,  $m = 0$  becomes a maximum along the planted ground state direction. From a computational perspective, we anticipate that a well-designed local algorithm will be likely to succeed below this temperature as it follows the free energy gradients leading it to the planted solution.

To further appreciate the first-order transition, we illustrate the discontinuity at  $T_c$  of some key thermodynamic quantities for a few representative  $\alpha$  values. A key observation is that the nature of the transition changes “gracefully” from first to second-order as  $\alpha$  increases past unity.

First, we recall (see, for example, Ref. [54]) that the ensemble partition function is related to the mean-field free energy density  $f(m)$  as

$$Z_N(\beta) \doteq \int_{-1}^1 \exp \left[ -N\beta f_\beta(m) \right] dm.$$

Using Laplace’s method, we obtain the “log partition

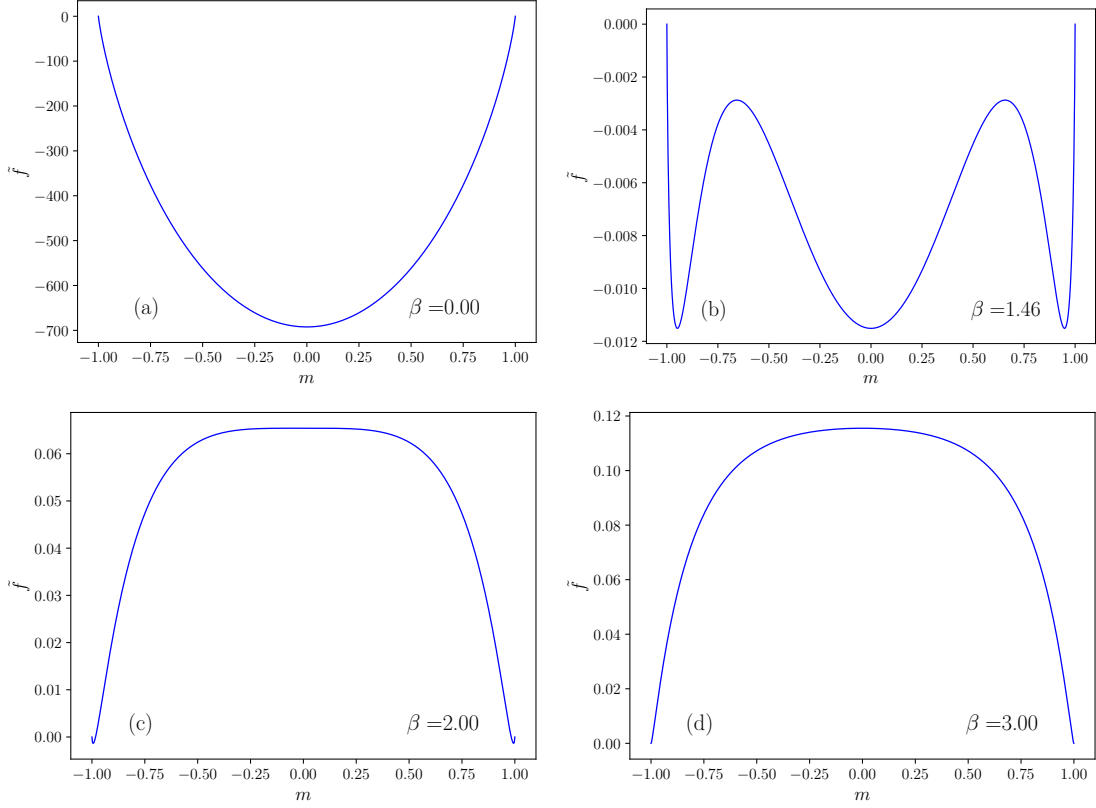


FIG. 4.  $\tilde{f}$  for  $\alpha = 1.5$ . Panel (a) shows the high temperature phase, whereas panel (b) shows the first-order phase transition at  $T_c = 0.68$ . The paramagnet at  $m = 0$  remains stable until  $T_u = 0.5$ , shown in panel (c), at which point it becomes unstable. For  $T < T_u$  [panel (d)], the ground states are easily attainable.

density”

$$\lim_{N \rightarrow \infty} \frac{1}{N} \log Z_N(\beta) = -\beta \min_{m \in [-1, 1]} f_\beta(m),$$

where an additive term arising from Laplace’s formula vanishes for large  $N$ . We define the maximum and maximizer of the free energy density, respectively, as

$$f^*(\beta) \triangleq \min_{m \in [-1, 1]} f_\beta(m)$$

and

$$m^*(\beta) \triangleq \arg \min_{m \in [-1, 1]} f_\beta(m).$$

The limiting internal energy density is

$$\begin{aligned} e(\beta) &\triangleq \lim_{N \rightarrow \infty} \frac{1}{N} E_N(\beta) \\ &= \lim_{N \rightarrow \infty} -\frac{1}{N} \frac{d}{d\beta} \log Z_N(\beta) \end{aligned}$$

so that

$$e(\beta) = \frac{d}{d\beta} \beta f^*(\beta).$$

In Fig. 5, we plot  $f^*(\beta)$  for four values of  $\alpha$ , with the corresponding  $\beta_c = 1/T_c$  located at the vertical dashed line. For the first two values of  $\alpha$  (0.5 and 0.75), the log partition function is clearly a smooth function on either side of the transition but is not differentiable at  $\beta_c$ . At  $\alpha = 2.5$ , the two sides still have different derivatives but

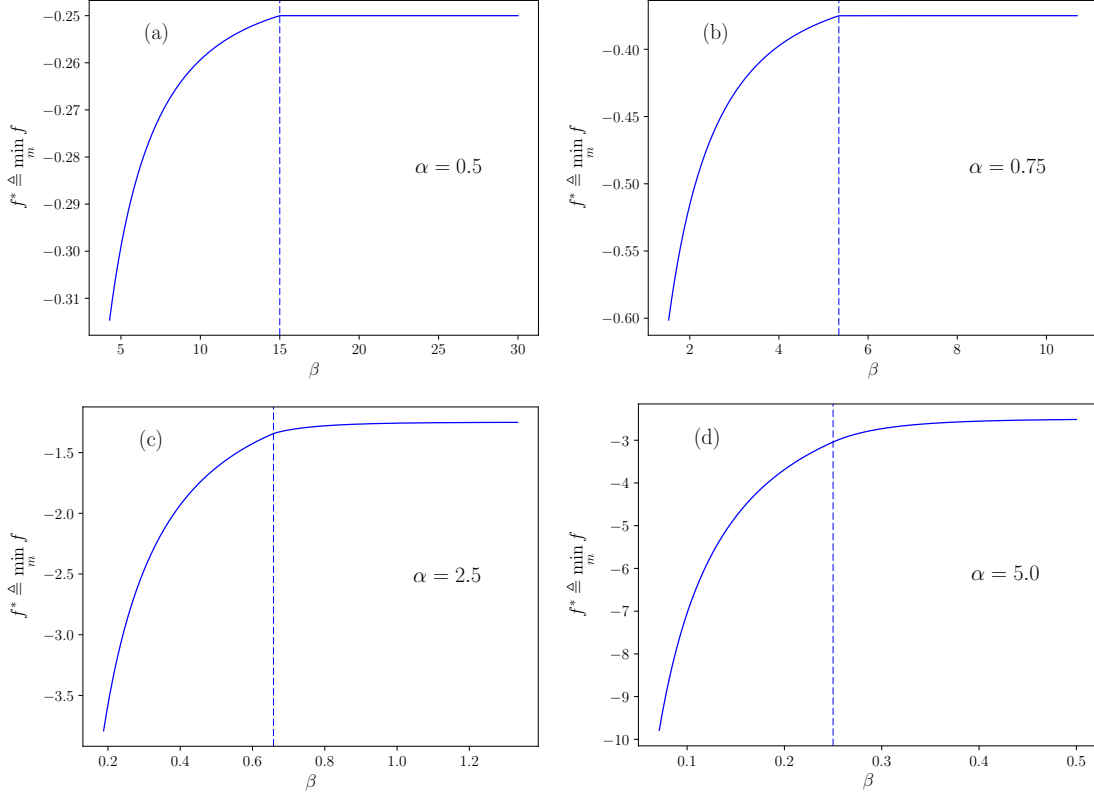


FIG. 5.  $f^*(\beta)$  corresponding to the asymptotic log partition function density for four values of  $\alpha$ . The first-order transition occurs at  $\beta_c = 1/T_c$  denoted by the dashed vertical line in each panel. When  $\alpha = 0.5$ , shown in panel (a), and  $\alpha = 0.75$ , shown in panel (b), the function is clearly not analytic at  $\beta_c$ . It remains so for  $\alpha = 2.5$  [panel (c)], but the discrepancy in the derivatives on the two sides has lessened. When  $\alpha = 5$  the function, plotted in panel (d) visually appears smooth at  $\beta_c$ .

the discrepancy is diminished, and when  $\alpha = 5$  it is no longer visibly discernible.

Figure 6 shows the magnetization  $m^*(\beta)$  for the same four values of  $\alpha$ ; the smaller  $\alpha$  values show a sharp, discontinuous bifurcation at  $\beta_c$  at which point the system abruptly shifts from being paramagnetic to being closely aligned with the planted solution. The jump is still apparent but attenuated for  $\alpha = 2.5$ , while for  $\alpha = 5$  the bifurcation starts to resemble a second-order ferromagnetic transition.

Finally, the internal energy densities are plotted in Fig. 7. Once again, the energies drop discontinuously at  $\beta_c$  from some excited values to the planted solution ground state energies (per spin) of  $-\alpha/2$ , which is most evident for the smaller two values of  $\alpha$ , diminished but still apparent when  $\alpha = 2.5$ , and essentially invisible for  $\alpha = 5$ . Note that the energy gap between

the two sides of  $\beta_c$  is larger when  $\alpha = 0.75$  than when  $\alpha = 0.5$ , in accordance with its higher transition temperature. The gap closes again for large  $\alpha$  but the energies thereafter progress continuously to their ground-state values.

The gradual reversion to a continuous transition is sensible in light of the construction procedure; when  $M$  is very large, the matrix  $\tilde{\mathbf{J}}$  converges to  $-\alpha\mathbf{\Sigma}$ , where  $\mathbf{\Sigma}$  is the  $\mathbf{w}$  covariance matrix. Since  $\Sigma_{ij} = -\frac{1}{N-1}$ , this implies that in the large  $\alpha$  limit, the system becomes a (scaled) Curie-Weiss ferromagnet with a transition temperature of  $T \sim \alpha$ .

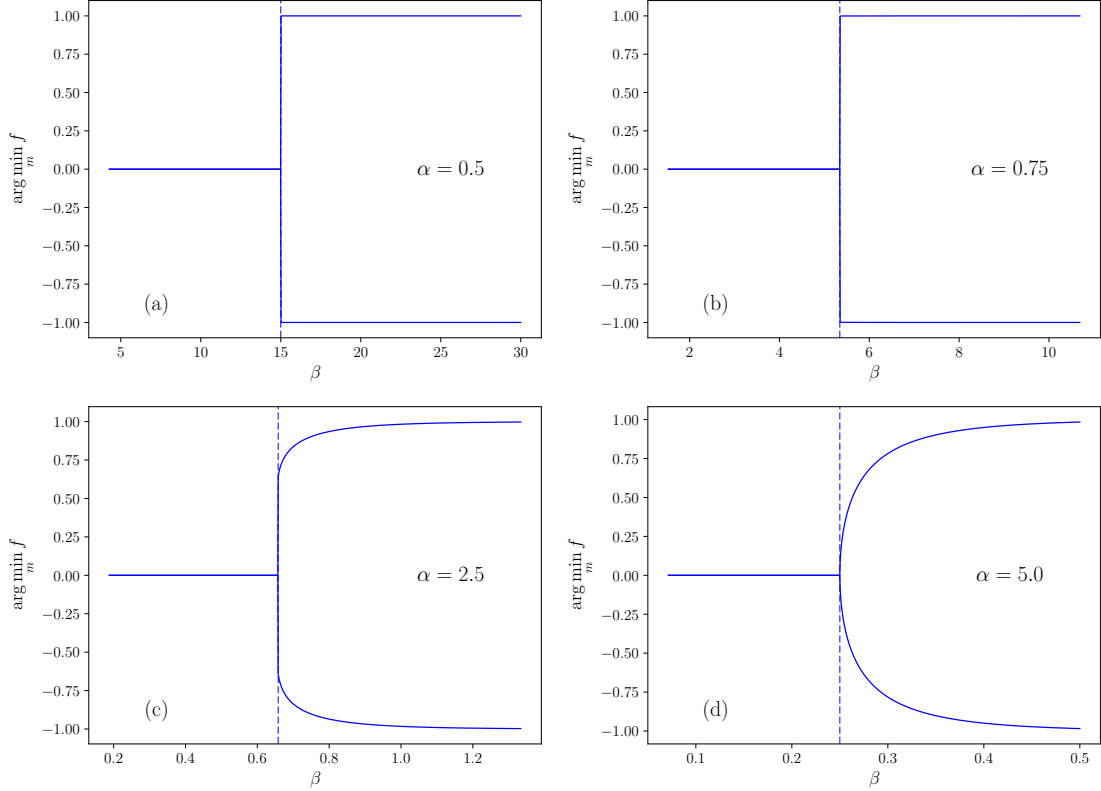


FIG. 6. Magnetizations  $m^*(\beta)$  for the same values of  $\alpha$  used in Fig. 5. The first-order transition appears as an abrupt bifurcation at  $\beta_c$  (dotted vertical lines) most apparent for  $\alpha = 0.5$  (a) and  $\alpha = 0.75$  (b). When  $\alpha = 2.5$  (c) the discontinuity is diminished, while for  $\alpha = 5$  (d) the bifurcation appears to have changed to second-order.

### C. Monte Carlo simulation

In this section we perform finite-temperature Monte Carlo simulation of fully-connected spin glasses with couplers drawn from a Wishart distribution. To detect the existence of a finite-temperature phase transition we measure the Binder cumulant [55] given by

$$g = \frac{1}{2} \left( 3 - \frac{[\langle m^4 \rangle_T]_{\text{av}}}{[\langle m^2 \rangle_T]^2_{\text{av}}} \right), \quad (21)$$

where  $\langle \dots \rangle_T$  represents a thermal average,  $[\dots]_{\text{av}}$  represents a disorder average, and

$$m = \frac{1}{N} \sum_{i=1}^N s_i \quad (22)$$

is the magnetization per spin. Although the model is disordered, it orders into a ferromagnetic phase because

the ferromagnetic ground state is planted by construction. In general, the Binder ratio scales as

$$g = \tilde{G} \left( N^{1/\nu} [T - T_c] \right), \quad (23)$$

where  $T$  is the temperature and  $T_c$  the critical temperature. In general,  $g(T = T_c)$  is independent of  $N$  and, as such, one can determine  $T_c$  by the point where data for different  $N$  cross. However, because as we shall see the transition is first order, the shape of the Binder ratio as a function of temperature is somewhat different to the commonly-known shape in second-order phase transitions (see, for example, Ref. [56]). When a first-order phase transition is present, the Binder ratio starts at  $g(T \rightarrow 0) \rightarrow 1$ , dips into  $g(T) < 0$  and then plateaus to  $g(T \rightarrow \infty) \rightarrow 0$ . One can determine the critical transition temperature by either extrapolating  $g(T_{\min}, L)$  for  $L \rightarrow \infty$ , where  $T_{\min}$  is the minimum value of  $g$ . Alter-



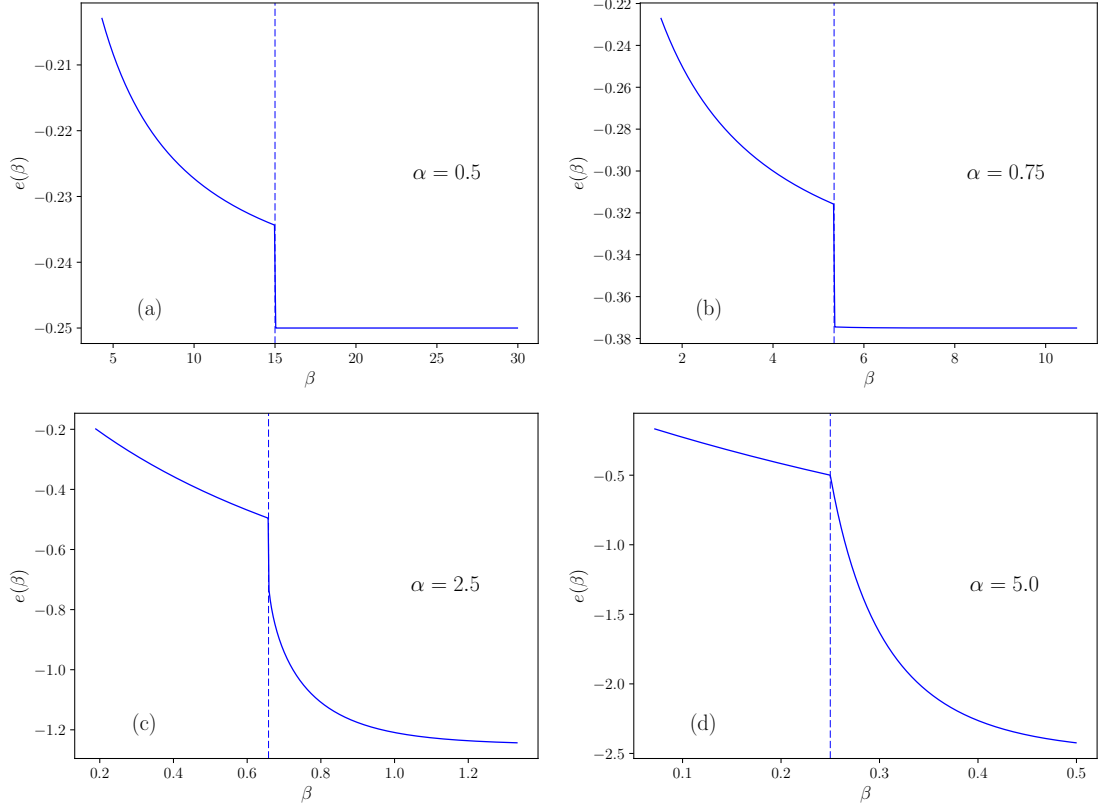


FIG. 7. Internal energy densities  $e(\beta)$  for the values of  $\alpha$  used in Figs. 5 and 6. For  $\alpha = 0.5$  (a) and  $\alpha = 0.75$  (b) the energy density drops discontinuously from an excited value to the planted solution ground state energy density of  $-\alpha/2$ . For  $\alpha = 5$  (d) the discontinuity appears to have disappeared and the internal energy drops gradually with temperature.

natively, one can study the crossing of data for different system sizes  $N$  in the range  $0 \leq T \leq T_{\min}$ . The former is usually easier to use to detect a transition, because the latter is often close to  $T = 0$  where data for different  $N$  do not splay enough to see a clean crossing. However, as shown in Refs. [57, 58],  $g(T_c) - g(T_{\min}, N) \sim 1/N$ , whereas the corrections to scaling in Eq. (23) decrease proportional to  $g(T_c) - g(T, N) \sim 1/N^2$ . As such, here we determine the position of the critical temperature by studying the crossing of the Binder ratio.

To further corroborate the existence of a first-order transition to a ferromagnetic phase we study the distribution  $P(m)$  of the magnetization [Eq. (22)]. Close to the transition temperature where latent heat is present the order parameter should signal two competing phases, i.e., peaks at  $|m| \rightarrow 1$ , as well as a competing peak at  $m = 0$ .

The simulations are done using parallel tempering Monte Carlo [16, 17]. Thermalization is verified by ensuring that all measured observables are independent of simulation time. We do this by analyzing how the results for all observables vary when the simulation time is successively increased by factors of 2 (logarithmic binning). We require that the last three results for all observables agree within error bars. Simulation parameters are shown in Tab. I. Error bars are determined via a jackknife analysis over the disorder.

Figure 8 shows data for the Binder ratio as a function of temperature. Figures 8(a), 8(c), and 8(e) show the Binder ratio for  $\alpha = 0.50, 0.75$ , and  $1.00$ , respectively, over the whole temperature range. A negative dip signaling a first-order transition is clearly visible. Figures 8(b), 8(d), and 8(f) zoom into  $T \in [0, T_{\min}]$ . As can be seen, the data cross for all three  $\alpha$  values. From the

TABLE I. Parameters of the simulations.  $N_{\text{sa}}$  is the number of samples,  $N_{\text{sw}} = 2^b$  is the total number of Monte Carlo sweeps for each of the  $N_T$  replicas for a single sample,  $T_{\text{min}}$  [ $T_{\text{max}}$ ] is the lowest [highest] temperature simulated, and  $N_T$  is the number of temperatures used in the parallel tempering scheme for each system size  $N$  and  $\alpha$ .

$\alpha$	$N$	$N_{\text{sa}}$	$b$	$T_{\text{min}}$	$T_{\text{max}}$	$N_T$
0.50	48	1000	22	0.0650	1.40000	150
0.50	64	1000	22	0.0650	1.40000	150
0.50	96	1000	26	0.0650	1.40000	150
0.75	48	1000	21	0.1150	1.40000	130
0.75	64	1000	21	0.1150	1.40000	130
0.75	96	1000	21	0.1150	1.40000	130
0.75	128	1000	21	0.1150	1.40000	130
1.00	48	1000	21	0.1150	1.40000	130
1.00	64	1000	21	0.1150	1.40000	130
1.00	96	1000	21	0.1150	1.40000	130
1.00	128	1000	21	0.1150	1.40000	130
1.00	192	1000	21	0.1150	1.40000	130

crossing points we estimate  $T_c(\alpha = 0.50) = 0.065(5)$ ,  $T_c(\alpha = 0.75) = 0.188(1)$ , and  $T_c(\alpha = 1.00) = 0.335(5)$ , in perfect agreement with our analytical estimates. Note that  $T_c(\alpha) \rightarrow 0$  for decreasing  $\alpha$ , i.e., the problems become increasingly harder numerically for smaller values of  $\alpha > \alpha_c$ .

In Fig. 9 data for the magnetization distribution are shown for  $\alpha = 0.50, 0.75$ , and  $1.00$  and  $T \approx T_c$ . In all three cases peaks at  $|m| \rightarrow 1$  are visible, signaling ferromagnetic order. However, a third peak for  $m = 0$  grows with increasing  $N$ , thus signaling a jump in the magnetization close to the transition. Note that for decreasing  $\alpha$  the first-order transition is more pronounced.

#### IV. HARDNESS PHASE TRANSITION

Having studied the phase behavior of the WPE and elicited properties that would plausibly correlate with algorithmic difficulty, we now proceed to examine some empirical results in which we probe the typical time to find the ground state using a parallel tempering (PT) Monte Carlo method. This algorithm is widely used to simulate complex physical and biological systems; contemporary implementations routinely examine spin glasses with several thousand variables and are quite competitive as optimization algorithms as well [59, 60]. In essence, it is a careful stochastic local search [61]

method such that configurations are allowed to escape metastable states by travelling to high temperatures and return afresh to probe the low-energy landscape while ensuring asymptotically correct equilibrium sampling at each temperature.

We seek to check and account for the existence of an algorithmic hardness peak, a sharp increase in the time required to find the ground state as  $\alpha$  is varied. Unfortunately, our numerical studies were severely hampered by the extreme difficulty of the WPE in its hard regime, which disallowed us from empirically localizing the peak for sizes that would otherwise be considered modest by the standards of other Ising ensembles (e.g., the SK model). We systematically reduced the sizes in which PT failed, within our computational resources, to approach the planted ground-state energy for a range of  $\alpha < 1$  values. Reliable statistics were finally obtained for  $N = 32$ ; the plot of median time to solve the problem, or more precisely and as justified in Sec. IV C, to find a state with energy within  $\epsilon = 10^{-7}$  of that of the planted ground state, is shown in Fig. 10. Details of the simulation protocol are found in Appendix VII F.

The hardness peak is evident at  $\alpha = 3/N = 3/32$ , where the problems typically took around 5 minutes to solve. As expected, the instances get steadily easier for the larger  $\alpha$  values, but the observation that  $M = 1$  is easy may initially seem surprising. As alluded to previously, this is due to the fully-frustrated nature of  $M = 1$ , giving rise to a tremendous number of low-energy states, coupled with the fact that standard double-precision floating point arithmetic was used for the computations. Consequently, the potential exists for several states that under unbounded precision would have had close but nonetheless distinct energies to be mapped to the same value. Under the low-energy degeneracy associated with  $M = 1$ , one thus observes a huge and exponentially increasing (in  $N$ ) number of numerically indistinguishable “ground states;” finding one such acceptable state turns out to be relatively easy [62].

Yet, the number of solutions alone cannot be expected to predict problem difficulty; in the computationally easy large- $M$  regime, the Hamiltonian starts to look increasingly like a ferromagnet, and so is essentially assured of having a unique solution. Thus, in contrast to uncorrelated problems like number partitioning in which an easy-hard transition is observed at the parameter value such that “perfect” solutions to the problem disappear [35, 42], the correlations resulting from the planting procedure give rise to another factor influencing difficulty and which leads to the observed easy-hard-easy

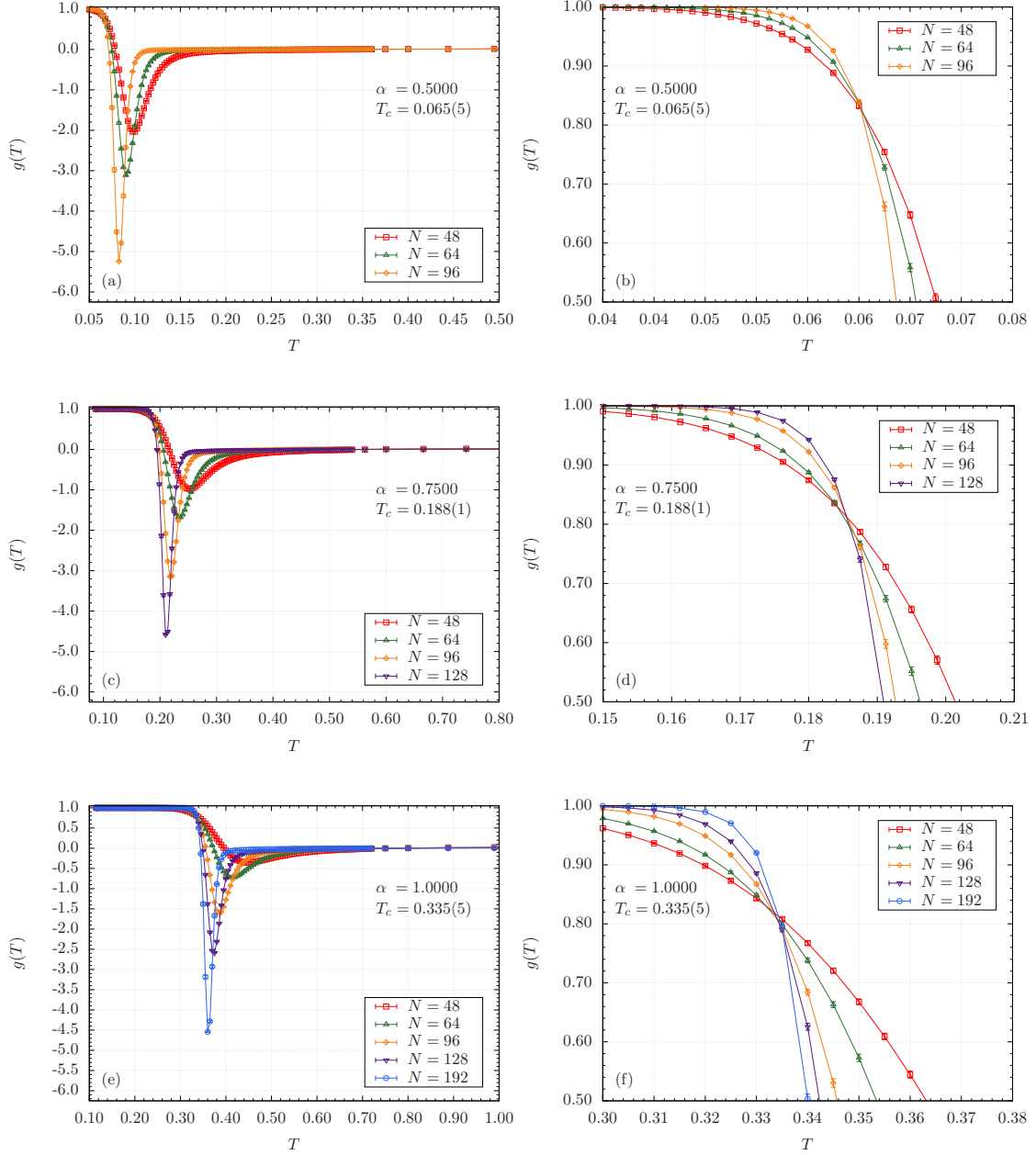


FIG. 8. Binder ratio as a function of temperature for  $\alpha = 0.50, 0.75$ , and  $1.00$ . Panels (a), (c), and (e) show data for the complete temperature range, whereas panels (b), (d), and (f) zoom into the region where  $T \in [0, T_{\min}]$ . A first-order transition is clearly visible.

pattern. We conjecture that hardness at a given  $\alpha$  is determined by two competing factors: First, the number of solutions satisfying the integer program and, second,

the *intrinsic* search space size of the integer program. It seems plausible that the regime in which the ratio of these two quantities is minimized signals the hard-

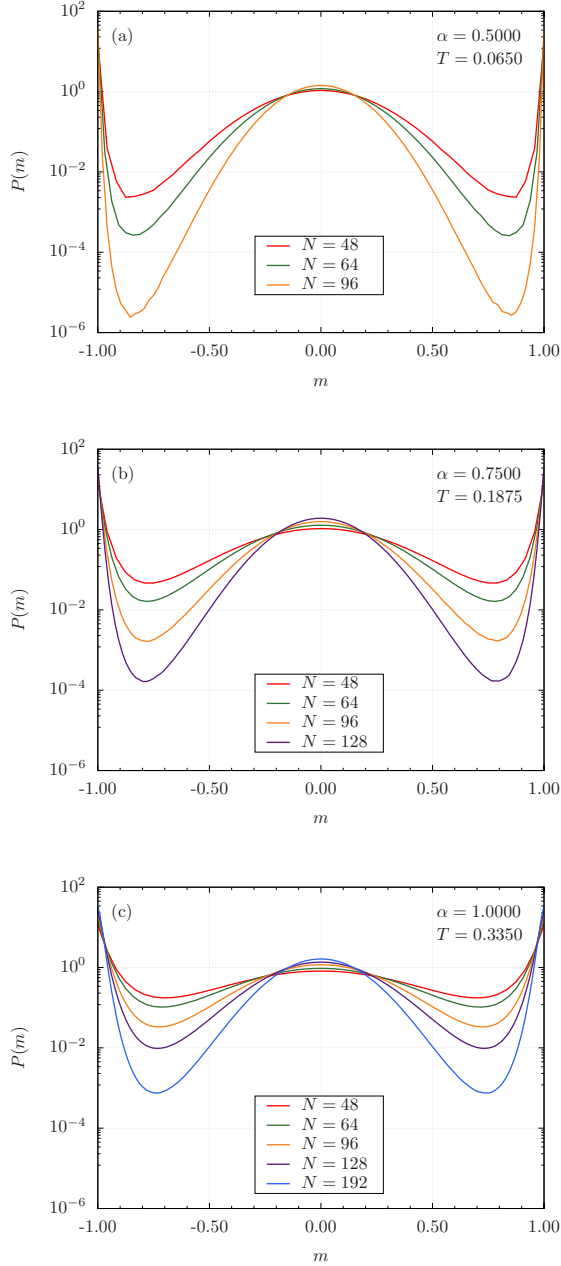


FIG. 9. Order parameter distributions for  $\alpha = 0.50, 0.75$ , and  $1.00$  in a linear-log scale. The data clearly show three peaks—two close to  $|m| \rightarrow 1$  and one close to  $m = 0$ —for temperatures close to the critical temperatures determined in Fig. 8. Close to the transition there are competing phases, i.e., a multi-value order parameter.

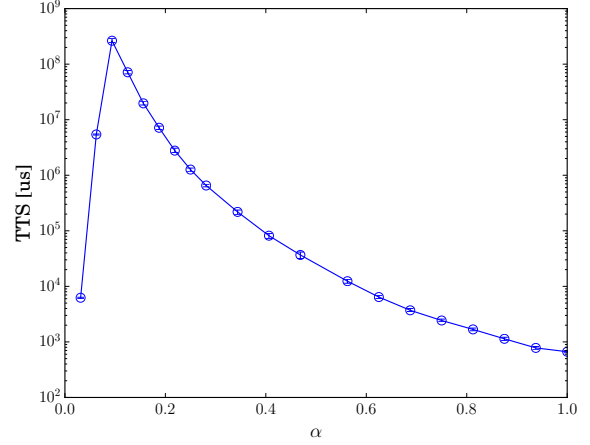


FIG. 10. Optimized parallel tempering Monte Carlo time-to-solution for the WPE at  $N = 32$  for 20 values of  $1 \leq M \leq N$ . This is a prototypical easy-hard-easy pattern with the peak occurring when  $M = 3$ . Note that the vertical axis is logarithmic and spans many orders of magnitude. As discussed in Sec. IV C, a “solution” in this case is defined as finding a state whose energy lies within  $\epsilon = 10^{-7}$  of the planted ground-state energy.

ness transition as this would be analogous to having the fewest needles in the biggest haystack. Our aim is to analytically predict the hardness peak for the  $N$ -variable WPE at given  $\alpha$  and numerical tolerance  $\epsilon$ .

To determine the number of satisfying solutions under precision  $\epsilon$ , we first derive the ensemble-averaged energy histogram for the WPE, discussed in Sec. IV A. We then formalize our notion of the intrinsic search space in Sec. IV B; as anticipated in Sec. II A, the size of the nullspace of  $\mathbf{W}^T$  plays a key role. Finally, we make our quantitative conjecture regarding the transition location in Sec. IV C and show that it precisely predicts the location of the peak in Fig. 10. In Sec. IV D, we present some preliminary numerically-obtained features of the WPE energy landscape to complement the main analysis performed here.

#### A. Energy histogram

We consider an ensemble of problems over  $N$  spins with parameter  $M$  such that

$$\tilde{\mathbf{J}} = -\frac{1}{N} \mathbf{W} \mathbf{W}^T$$

with the columns  $\{\mathbf{w}^\mu\}$  of  $\mathbf{W}$  simulated as  $\mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$  and Hamiltonian

$$H(\mathbf{s}) = -\frac{1}{2} \mathbf{s}^T \tilde{\mathbf{J}} \mathbf{s}$$

for  $\mathbf{s} \in \mathbb{S}^N$ . Note that the diagonal elements are included in this formulation as it is presently more convenient to take the ground-state energy to be zero. Without loss of generality, we assume the ferromagnetic ground state is planted so that for all  $\mu$

$$\sum_{i=1}^N w_i^\mu = 0$$

. We seek the distribution of energies marginalized over the problem ensemble

$$p_E(e) \triangleq \int_{\{\mathbf{w}^\mu\}} p_E(e|\{\mathbf{w}^\mu\}) f(\mathbf{w}^1, \dots, \mathbf{w}^M) d\{\mathbf{w}^\mu\}, \quad (24)$$

where  $p_E(e|\{\mathbf{w}^\mu\})$  is the probability of drawing by *uniform* sampling a state with energy  $E = e$  from the given problem.

The result is surprisingly simple; the energies follow a gamma distribution with density

$$p_E(e) = \begin{cases} \frac{1}{\Gamma(M/2)} e^{M/2-1} \exp(-e) & \text{for } e \geq 0 \\ 0 & \text{for } e < 0 \end{cases} \quad (25)$$

and cumulative distribution  $P_E(e) \triangleq \Pr(E \leq e)$

$$P_E(e) = \begin{cases} \frac{1}{\Gamma(M/2)} \gamma\left(\frac{M}{2}, e\right) & \text{for } e \geq 0 \\ 0 & \text{for } e < 0 \end{cases}, \quad (26)$$

where  $\Gamma(x)$  and  $\gamma(x, y)$  are the gamma and incomplete gamma functions [63], respectively. Figure 11 displays  $p_E(e)$  for the WPE when  $N = 64$  for a few values of  $M$ . When  $M = 1$ , the density is overwhelmingly concentrated on low-energy values; combined with bounded precision, this degeneracy is responsible for the problems being easy. By the properties of the gamma distribution, the mean and standard deviation of  $E$  are  $M/2$  and  $\sqrt{M/2}$  respectively; consequently (nonplanted) low-energy states become exponentially less likely as  $M$  increases.

The calculation of the distribution is shown in Appendix VII C. It is quite similar to Mertens' [30] analysis of the number partitioning problem cost density, but here the energy is a more complicated quadratic function with a

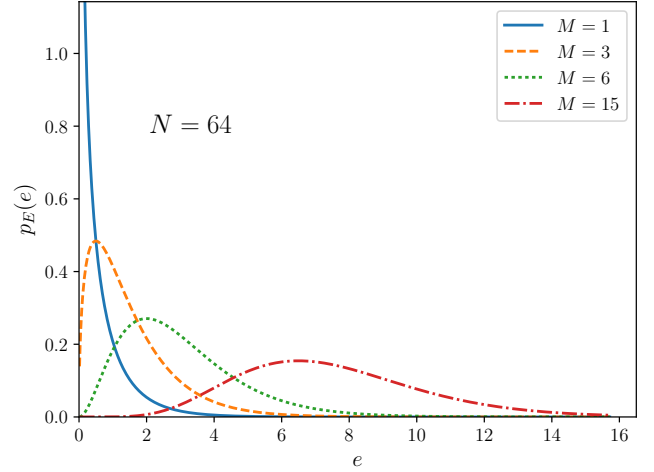


FIG. 11. The energy distribution  $p_E(e)$  of a WPE instance of size  $N = 64$  at different values of  $M$ . When  $M = 1$  (solid blue line), the density is overwhelmingly concentrated on values very close to the ground-state energy of  $E = 0$ . The energy mean and standard deviation (fluctuations) are  $M/2$  and  $\sqrt{M/2}$  respectively; when  $M = 15$  (dash-dotted red line) there is vanishing probability of drawing a low-energy state by random chance.

correlated coupler matrix rather than a rank-1 absolute value discrepancy over independent coefficients.

If the coupler matrix  $\mathbf{J}$  rather than  $\tilde{\mathbf{J}}$  were used, the density would simply be translated by the ground state energy of  $-\alpha N/2$ , i.e.,

$$p_E(e) \leftarrow p_E(e + \alpha N/2).$$

## B. Intrinsic search space cardinality

We have speculated that because we aim to solve an Ising-constrained linear program or equivalently, to solve for a state  $\mathbf{s}^* \in \mathbb{S}^N$  lying in the nullspace of  $\mathbf{W}^T$ , the dimensionality of the nullspace plays some role in problem complexity. In this section, we clarify the notion and extract an  $M$ -dependent intrinsic search space. This is defined as a discrete set of reduced-dimensional states over which one would need to exhaustively search to solve the problem without reference to any information about the values of  $\mathbf{W}$ . In some sense, this is analogous to the reduced complexity enjoyed by various optimization problems on graphs of low treewidth. The result is that when  $\mathbf{W}^T$  has  $M$  independent rows, one only needs to

consider  $O(2^{N-M-1})$  states in brute force search; the subtraction by one is simply to remind that the problem is invariant to global spin flips. We note that the following procedure is not meant for literal implementation because we are disregarding, for example, numerical issues that may be important in practice, but simply to show that up to a polynomial prefactor the exhaustive search complexity is exponential in  $K$ , where

$$K \triangleq \dim(\text{null}(\mathbf{W}^T)) = N - M.$$

We remind the reader that the integer programs are always feasible due to the planting construction. Let the vectors

$$\mathcal{V} \triangleq \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_K\}$$

span  $\text{null}(\mathbf{W}^T)$ . Such a basis can be obtained in polynomial time using, for example, the singular value decomposition of  $\mathbf{W}$ . The integer program can be equivalently phrased as

$$\begin{aligned} &\text{find } \mathbf{s}^* \in \mathbb{S}^N \\ &\text{such that } \mathbf{V}\mathbf{x} = \mathbf{s}^* \text{ is consistent,} \end{aligned}$$

where  $\mathbf{V}$  is the  $N \times K$  matrix whose columns are  $\{\mathbf{v}_i\}$  and  $\mathbf{x} \in \mathbb{R}^K$ . For any state  $\mathbf{s}$ , one can in polynomial time either solve for satisfying coefficients  $\mathbf{x}$  or show that no such coefficients exist using standard linear algebra techniques. While a solution will eventually be found when the method is repeated for all possible states, we now show that not all  $\mathbf{s} \in \mathbb{S}^N$  need be checked.

We first obtain a  $K \times K$  matrix  $\tilde{\mathbf{V}}$  comprised of a subset of  $K$  independent rows from  $\mathbf{V}$ . Such a matrix can be determined using for example the QR decomposition with pivoting [64]. We apply this decomposition to the columns of  $\mathbf{V}^T$ :

$$\mathbf{V}^T \mathbf{P} = \mathbf{Q} \mathbf{R},$$

where  $\mathbf{P}$  is an  $N \times N$  column-permuting matrix,  $\mathbf{Q}$  is  $K \times K$  and orthogonal, and  $\mathbf{R}$  is  $K \times N$  and upper-triangular. This ensures that the first  $K$  columns of  $\mathbf{V}^T \mathbf{P}$  are independent, so we take

$$\tilde{\mathbf{V}} = \mathbf{P}_K^T \mathbf{V},$$

where  $\mathbf{P}_K$  is an  $N \times K$  submatrix comprised of the first  $K$  columns of  $\mathbf{P}$  and encodes which rows of  $\mathbf{V}$  were selected for  $\tilde{\mathbf{V}}$ . These steps are computed only once for a given problem.

Now let  $\tilde{\mathbf{s}}$  be an Ising state of length  $K$ . The system

$$\tilde{\mathbf{V}} \tilde{\mathbf{x}} = \tilde{\mathbf{s}}$$

with respect to partial state  $\tilde{\mathbf{s}}$  always has a solution because  $\tilde{\mathbf{V}}$  is full-rank. After solving for  $\tilde{\mathbf{x}}$ , we verify whether this defines a feasible solution over *all* Ising variables by computing

$$\mathbf{y} = \mathbf{P}^T \mathbf{V} \tilde{\mathbf{x}}.$$

The first  $K$  components of  $\mathbf{y}$  will be  $\tilde{\mathbf{s}}$  by construction; if the rest are also in  $\{\pm 1\}$  then  $\mathbf{y}$  is a feasible solution. If not, we choose another  $\tilde{\mathbf{s}}$  and repeat the process starting with solving for  $\tilde{\mathbf{x}}$ . This shows that up to the polynomial overhead involved in the various steps, one only needs to consider the number of states in  $\mathbb{S}^K$ ; in fact due to the spin-flip symmetry, the intrinsic search space contains  $2^{K-1}$  elements. In the case where  $M \geq N$  and so the nullspace is one-dimensional, only a single state need be checked.

### C. Expression for the hardness transition

We now synthesize the two preceding antithetical factors into an expression that predicts when problems are likely to become difficult. We first heuristically obtain an approximation to the number of ground states under a discretized parameter representation using the continuous energy histogram  $p_E(e)$ .

At a given problem size, finite precision means we cannot represent all the “true” values of the Gaussian variables  $\{\mathbf{z}^\mu\}$  used to generate  $\mathbf{W}$  nor can we in general exactly represent the values of  $\mathbf{W}$  resulting from the planting procedure using the rounded  $\{\mathbf{z}\}$  variables or even the exact energy of a state relative to the rounded  $\mathbf{W}$ . Rounding and discretization cause errors in the representation of some energies; in particular, states that had distinct energies under arbitrary precision may be mapped to numerically indistinguishable values. Let  $\hat{E}$  be the random variable resulting from finite-precision representation of some true continuous-valued energy  $E$ ; its corresponding CDF  $P_{\hat{E}}(\epsilon) \triangleq \Pr(\hat{E} \leq \epsilon)$  now has step discontinuities. If we interpret the mapping of a state’s exact energy  $E$  to its representable approximation  $\hat{E}$  as a pseudo-random walk process leaving invariant the probability of having energy of at most  $\epsilon$ , then provided  $\epsilon$  is large enough to be representable we can approximate

the finite-precision energy CDF with the continuous distribution introduced in Sec. IV A:

$$\Pr(\widehat{E} \leq \epsilon) \approx \int_0^\epsilon p_E(e) de = P_E(\epsilon),$$

where the latter equality follows from the nonnegativity of  $E$ .

The value of  $\epsilon$  is in principle arbitrary but must be large enough to reflect the approximation error of the planted ground-state energy. We observe that under the aggregated discretization effects of the construction procedure, the true planted ground-state energy of zero is typically distorted to a value on the order of the so-called machine epsilon  $\approx 10^{-16}$  for double-precision arithmetic; to conservatively avoid missing planted ground states, we can define a successful solution to be one with energy at most  $\epsilon = 10^{-7}$  which avoids the possibility of error accumulation in the energy computed by the solver. We emphasize that larger values of  $\epsilon$  are certainly sensible; as will soon be shown, proceeding with our analysis with such values yields accurate predictions of the hardness transition for the important task of approximate (or relaxed) solution to the problem.

Recalling next the spin-flip symmetry of the Hamiltonian and the fact that the energy distribution was dominated in Laplace's approximation by the effect of paramagnetic states ( $m = 0$ ) which are far from the planted solution, we arrive at an expression of the ensemble-averaged expected number of solutions with observed energy of at most  $\epsilon$ :

$$\begin{aligned} \mathbb{E}[\#\{\widehat{E} \leq \epsilon\}] &= 1 + (2^{N-1} - 1)P_E(\epsilon) \\ &= 1 + (2^{N-1} - 1) \frac{1}{\Gamma(M/2)} \gamma\left(\frac{M}{2}, \epsilon\right). \end{aligned} \quad (27)$$

The leading “1” represents the persistence of the planted solution; its presence in the expectation may initially seem strange and for small values of  $\alpha$  it is indeed irrelevant, but as  $\alpha$  increases, the paramagnetically-contributed solutions start to vanish and its influence becomes dominant. Finally, we define the function trading off number of solutions to effective search space size

$$Q_N(M) \triangleq \frac{\mathbb{E}[\#\{\widehat{E} \leq \epsilon\}]}{2^{N-M-1}} \quad (28)$$

and anticipate that problems become most difficult at

$$M^* = \arg \min_{M \in \{1, \dots, N\}} Q_N(M)$$

as this coincides with the fewest number of solutions relative to the effectively constrained search space. In Fig. 12, we plot  $\log Q_N(M)$  for system sizes  $N = 32$  and  $N = 128$ , when  $\epsilon \approx 10^{-7}$  in accordance with our conservative rule for double-precision arithmetic. When  $N = 32$ ,  $M^* = 3$  in perfect accordance with the location of the hardness peak shown in Fig. 10. The latter size was far too large to solve in the hard regime within our constraints, but the plot suggests that to obtain the most difficult problems for this larger size,  $M$  needs to increase. It is instructive to predict the required scaling of  $M$  with  $N$  for maximally hard problems at this level of precision.

Figure 13 shows  $M^*$  as a function of  $N$ , where a clear linear scaling is apparent. By linear regression, we find this relation to be

$$M^* \approx 1.63 + 0.073N,$$

suggesting that the hardest problems at  $\epsilon = 10^{-7}$  occur when  $\alpha \approx 0.073$ . The general message is clear: to have truly difficult problems under precision constraints, the number of equations  $M$  in the integer program cannot be constant.

Note that the same analysis can be repeated for the Rademacher-discretized WPE introduced in Sec. II C, but in that case as we assume that energies can be computed exactly, rather than remain fixed, the tolerance  $\epsilon$  should now be the smallest attainable excited state energy, which is  $O(1/N^3)$ .

The reader may notice that in the definition of  $Q_N(M)$ , the count in the numerator used an  $\epsilon$ -relaxed definition of “solution” while the denominator search space cardinality was computed relative to the task of solving the problem exactly. One may indeed consider relaxing the integer program so that an energy of at most  $\epsilon \geq 0$  is sought as follows

$$\begin{aligned} &\text{solve } \mathbf{y} = \mathbf{W}^T \mathbf{s} \text{ such that} \\ &\frac{1}{2N} \|\mathbf{y}\|^2 \leq \epsilon \\ &\mathbf{s} \in \mathbb{S}^N. \end{aligned}$$

For general values of  $\epsilon$  however, this cannot be expected to fundamentally lower the exhaustive search complexity. It can be verified that this more permissive problem, which of course reduces to the exact case when  $\epsilon = 0$ , now allows searching for a satisfying  $\mathbf{s}$  in an ellipsoidal region enclosing null( $\mathbf{W}^T$ ) rather than being restricted to the nullspace proper. The search space volume, however, is still on the order of  $2^{K-1}$ . To justify this assertion, we

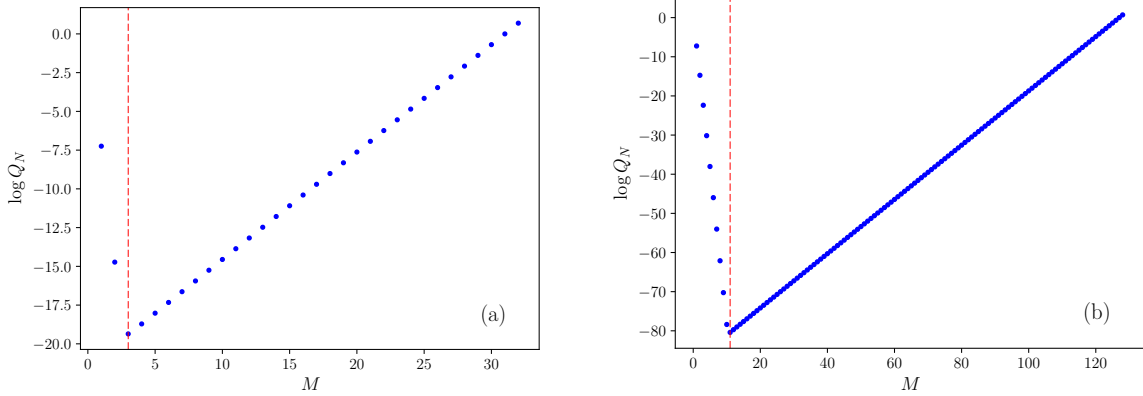


FIG. 12.  $\log Q_N(M)$  whose minimum, denoted by the dashed red line, is conjectured to predict the WPE hardness transition at precision  $\epsilon = 10^{-7}$  for sizes  $N = 32$  (left) and  $N = 128$  (right). For  $N = 32$ , the minimizer occurring at  $M = 3$  coincides with the maximum parallel tempering time-to-solution observed in Fig. 10. When  $N = 128$ , maximum difficulty is predicted to occur when  $M = 11$ , suggesting that at constant precision, generating the hardest problems requires scaling  $M$  with  $N$ .

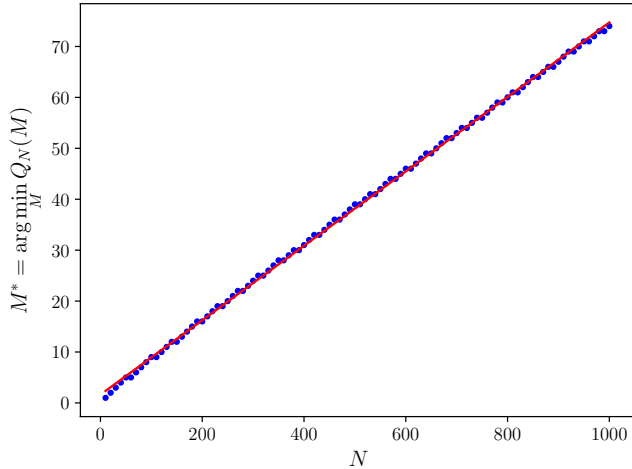


FIG. 13. Predicted scaling of  $M$  with  $N$  to maximize WPE hardness at precision  $\epsilon = 10^{-7}$ . True values of  $M^* = \arg \min_M Q_N(M)$  (blue points) clearly follow a linear relation. Regression approximates this relation as  $M^* \approx 1.63 + 0.073N$  (red line), suggesting that the hardest problems occur for this precision restriction at  $\alpha \approx 0.073$ .

now show that  $Q_N(M)$  can serve to localize the hardness peak for more general values of  $\epsilon$ .

Figure 14 shows the PT median times to solution for the same  $N = 32$  WPE ensemble considered so far, but for three values of  $\epsilon$  defining an acceptable solution:

$\epsilon = 10^{-7}$ ,  $10^{-5}$ , and  $10^{-3}$ . Unsurprisingly, the typical solution times decrease as the energy criterion becomes more permissive. Additionally, we note that the location of the hardness peak shifts to larger values of  $M$  as  $\epsilon$  is relaxed. Most importantly, in Fig. 15 we observe the predictive power of  $Q_N(M)$  at these values of  $\epsilon$ : in all three cases, the minimizer  $M^*$  of  $Q_N(M)$  at the respective values of  $\epsilon$  precisely corresponds with empirically-observed PT solution time. We have hence proposed a robust, theoretically-motivated framework for generating tunably-difficult problems over a wide range of approximate solution targets.

#### D. Properties of locally optimal states

As we have seen, the energy histogram derived in Sec. IV A has been useful in predicting the algorithmic properties of WPE instances. Nonetheless, this distribution does not provide information about topological aspects of the local minima, i.e., states that are energetically stable with respect to a single spin flip.

In this section, we briefly probe the properties of local minima using exhaustive search on small instances; we save analytic examination of these properties, along the lines of Bray and Moore's [65] analysis for the SK model, for later work.

A natural statistic to analyze is the expected number of local minima as  $\alpha$  is varied. Furthermore, it is instructive to define a residual energy histogram restricted to



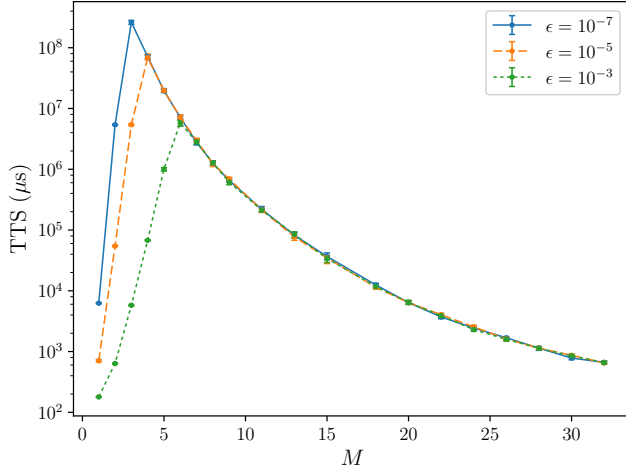


FIG. 14. Optimized parallel tempering time to approximate solution for the  $N = 32$  WPE as  $\epsilon$ , the acceptable excess energy over that of the ground state, is varied. As expected, the peak difficulty decreases as  $\epsilon$  increases, as this corresponds to making the objective more permissive. Interestingly, the hardness peak occurs at larger values of  $M$  as  $\epsilon$  grows. This phenomenon is discussed in the text, where we make analytical predictions of the peak location by reference to the function  $Q_N(M)$ , and illustrated in Fig. 15.

stable states. These are shown for a system with  $N = 24$  variables in Figs. 16 and 17, respectively. As expected, the number of minima decreases monotonically in  $\alpha$  as the ensemble tends towards a ferromagnet. At small  $\alpha$ , we observe a large number of minima, but also that the residual energy itself is likely to be small. This is consistent with the observation that at small  $\alpha$ , the problems with restricted precision are easy.

To further illustrate properties of the WPE, Fig. 18 shows examples of disconnectivity graphs for four specific instances at their respective  $M$ -values. Disconnectivity graphs are two-dimensional representations of high-dimensional energy landscapes. In their simplest form they depict the minima of the system and the lowest energy barrier connecting any two minima, where an energy barrier is defined as the highest energy value encountered along a specific pathway. The barriers represent the minimum increase in energy necessary to transition from one minimum to another. In this work, the minima of the system were obtained via complete enumeration and the barriers were calculated using a search over all possible pathways identified with flip-

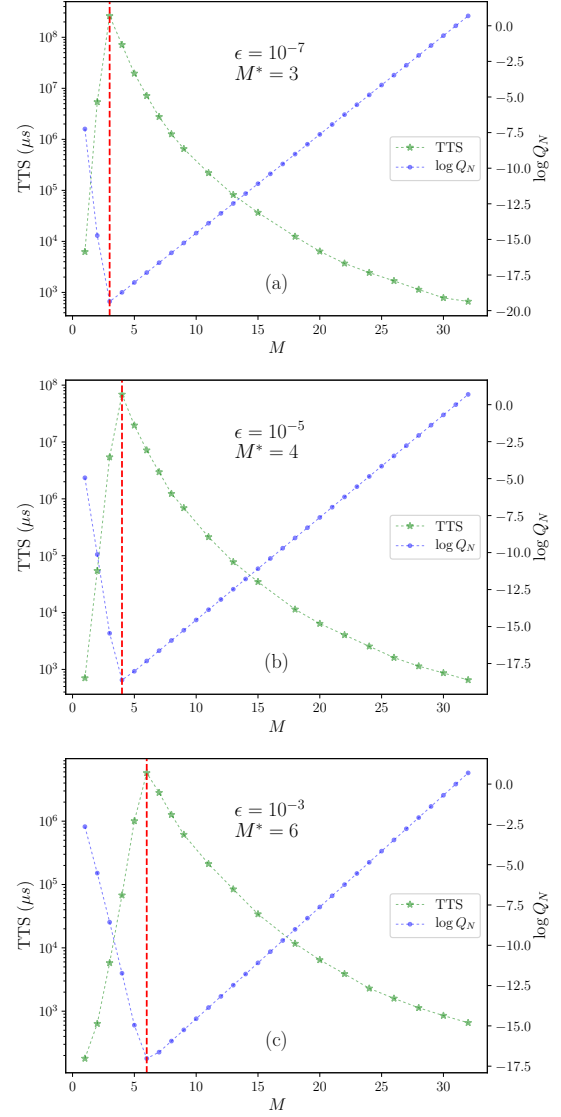


FIG. 15. Illustration of the predictive power of  $M^* = \arg \min_M Q_N(M)$  in localizing the algorithmic hardness peak for a range of target  $\epsilon$  values: (a)  $\epsilon = 10^{-7}$ ; (b)  $\epsilon = 10^{-5}$ ; (c)  $\epsilon = 10^{-3}$ . In each panel, we plot the parallel tempering solution times (green, left y-axis) for the  $N = 32$  WPE and  $\log Q_N(M)$  (blue, right y-axis) for the corresponding  $\epsilon$  values. The dashed red line displays the value of  $M^*$ , which agrees perfectly with the empirically-measured hardness peak.

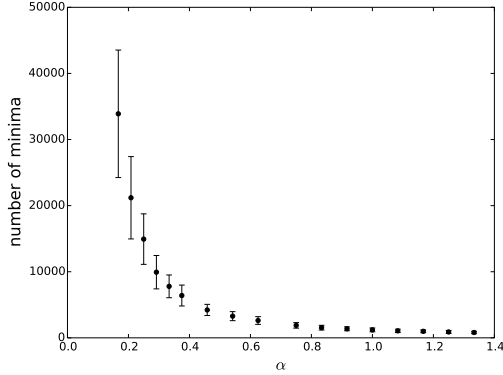


FIG. 16. Expected number of local optima for WPE instances of size  $N = 24$  as a function of  $\alpha$ .

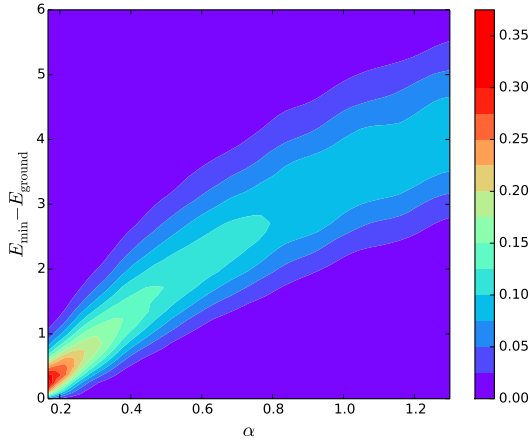


FIG. 17. Residual locally optimal energy distributions for WPE instances of size  $N = 24$  as a function of  $\alpha$ . For small  $\alpha$ , the distributions are concentrated on low residual energy values.

ping spins that are misaligned between pairs of minima. This method was outlined by Garstecki, Hoang and Cieplak [66]. To deal with the large number of minima in a computationally efficient manner, we further made use of an approximation based on the relative proximity of minima; specifically, for each minimum only the barriers to the 50 minima closest to it in Hamming distance were obtained. This approximation is based on the fact that transitions between two minima can also happen through basins of intermediate minima, and hence if

two minima are separated by a large Hamming distance it is likely that the lowest energy barrier between these two minima will be already represented via transitions between intermediate minima and their corresponding barriers.

In Fig. 18 the minima are represented by vertical bars whose lowest points denote their energies. On top, they are connected by lines converging to a common point representing the height of the barrier that needs to be crossed in order to transition between the connected minima. Due to the continuous nature of the energy values, we sort the minima into a hierarchical cluster structure whose endpoints comprise the intervals  $[E_b - \Delta\ell, E_b + \Delta\ell)$ , where  $E_b$  is the energy of the barrier and  $2\Delta\ell$  denotes the length of the interval. Minima whose connecting barriers fall within the same energy interval are sorted into a common cluster. Within an individual cluster the minima are arranged based on the number of spins in the up or corresponding down states. Minima which have a high number of up-state spins are sorted towards the left and, correspondingly, minima with a high number of down-state spins are sorted towards the right. Note, that this order strictly only applies within the individual cluster; the order of the individual clusters relative to each other is determined by the hierarchical structure. In this work we set  $\Delta\ell = 0.075$ . Figure 18 shows a clear progression in the energy landscape for which small values of  $M$ , e.g.,  $M = 4$  in panel (a), are characterized by a very large number of almost degenerate metastable states, and larger values of  $M$  [panels (b) – (d)] tend to break the degeneracy, emphasize the planted ground state, and make the landscape more funnel-like.

Figure 19 shows the distribution of the average barrier height within the first Hamming distance of its individual minima states. It represents the distribution of the average increase in energy necessary for the system to escape its minima via the shortest route, i.e., to transition to adjacent minima. For a given instance, the average  $\langle \dots \rangle$  was taken over the barriers to the minima which lie within the shortest Hamming distance to a given minimum, say minimum  $i$ .  $E_b$  are the energies of the individual barriers and  $E_{\min}$  is the energy of minimum  $i$ . The overall distribution is then obtained by sampling over all minima  $i$  of  $N = 100$  sample systems for each of the values of  $\alpha$ .

As can be seen in Fig. 19, at small values of  $\alpha$  the distribution is more spread than at large values, where additionally it is dominated by relatively small values of the average energy barrier. From an energy landscape

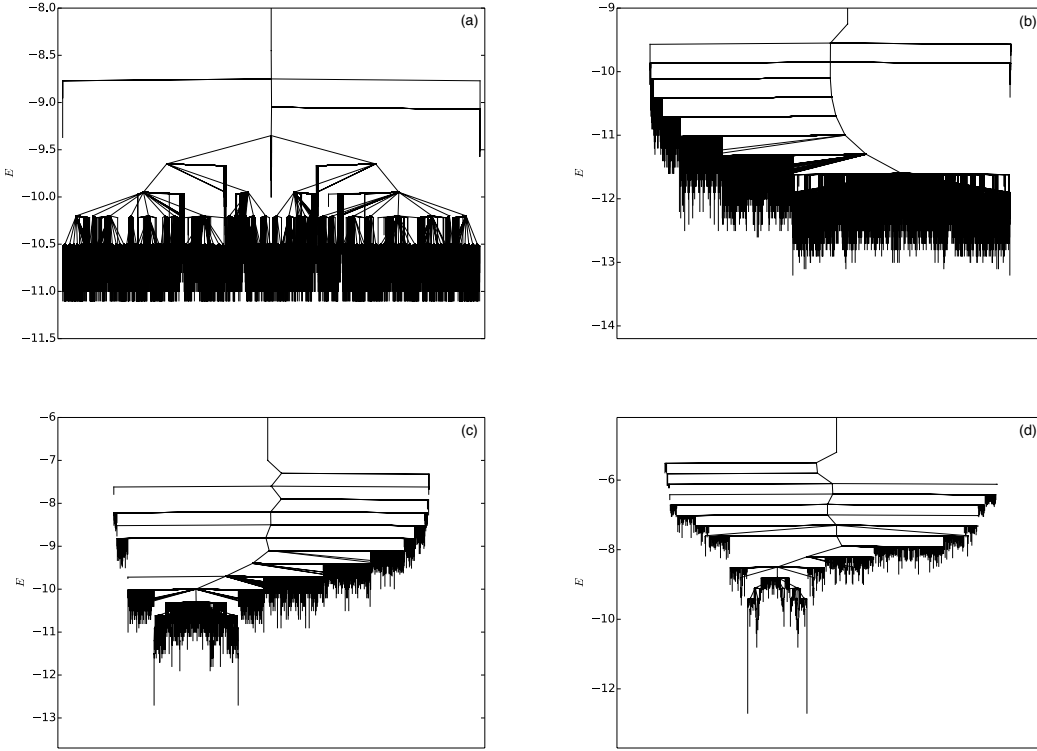


FIG. 18. Disconnectivity graphs (described in the text) representing the energy landscape of four specific  $N = 24$  WPE instances at various values of  $M$ . Leaf nodes of the tree structures depict local minima, and internal nodes represent barrier states. (a) When  $M = 4$  (top-left), we observe a tremendous near-degeneracy of a large number of metastable states with energy very close to that of the planted solution. (b) When  $M = 8$  (top-right), the degeneracy begins to lift, and the number of minima starts to decrease. As  $M$  increases to 15 (c) and 32 (d), the ground state becomes increasingly dominant and the problems computationally easier.

perspective this indicates that transitions from the minima at large  $\alpha$  are more likely to be achievable with less energy cost and therefore might be more probable than at small values of  $\alpha$  [67].

## V. DISCUSSION

We have proposed a planted Ising ensemble with several noteworthy physical and algorithmic properties. The model exhibits a first-order temperature transition, a persistent locally-stable paramagnetic state when  $\alpha < 1$ , and when represented with finite precision, an easy-hard-easy algorithmic difficulty profile. Its physical properties are consistent with the observed hardness of finding its

ground state; moderately-sized problems are extremely difficult in the hard regime. This meshes well with the intuition that the transition and paramagnetic stability give rise to a golf-course-like energy landscape. After deriving the instance-averaged energy distribution—which turns out to follow a gamma law—we compare the expected number of states matching a solution criterion with a quantity we introduced quantifying the intrinsic search space size at given  $\alpha$  to analytically predict the location of the hardness peak. The prediction is validated using solution times obtained with a highly-optimized implementation of parallel tempering Monte Carlo.

The first-order transition between the planted and paramagnetic phases is furthermore established by the TAP

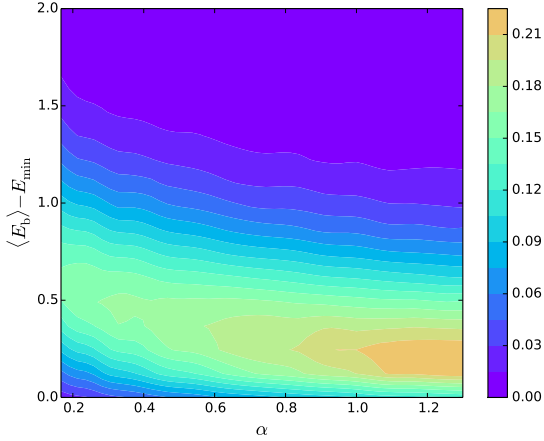


FIG. 19. Average barrier height distribution (described in the text) for the WPE ( $N = 24$ ). This quantity represents the average increase in energy incurred in transitioning from one minimum to a closely-lying neighbor. The distribution is more dispersed at small  $\alpha$ , where the barrier heights are furthermore larger in expectation than at larger  $\alpha$ . This indicates that transitions from the minima at large  $\alpha$  are more likely to be achievable with less energy cost and therefore might be more probable than at small values of  $\alpha$ .

analysis in Sec. III A, with alternative derivations supported by the replica method (Appendix VII E) and annealed approximations (Appendix VII D). Careful Monte Carlo simulations demonstrate the correctness of our analytically-predicted transition temperatures.

A connection is made with the anti-Hopfield model and this is developed analytically in Appendix VII E [36]. This analysis indicates that we might expect to see the impact of replica symmetry breaking if we focus on the region *orthogonal* to the planting space. In this paper we have not focused on this for two reasons. First, the emphasis has been on intermediate scale problems, where it is difficult to numerically establish such phenomena, finite size effects may dominate, and practical issues such as finite-precision representation may have greater impact. Second, at large  $N$  replica symmetry breaking has little impact on the free-energy barrier separating the bulk of the space from the planted solution, which is the primary driver of hardness for heuristic optimization in the planted case.

Nevertheless, we believe the qualitative description of the space orthogonal to the planted solution, that of a rough energy landscape with deep solutions almost or-

thogonal to the planted solution, is in effect. Roughness is apparent at the small scales we have worked with empirically. With this in mind, a modification of the planting procedure is pursued in Appendices VII D and VII E whereby the planted solution is partially penalized. Its leading-order energy is then tunable at fixed  $\alpha$ , at the cost of losing strict guarantees that it is a ground state. This modification, it is believed, might be the basis for interesting tests of dynamics (escape into or out of the planted solution). In particular, an interesting suggestion has been made that for this type of problem in the presence of a transverse field, quantum dynamics may be differentiated from classical counterparts [27]. The features of our model with this modified planting procedure represent a practical realization of many of the abstract model features underlying the population transfer hypothesis.

Future work will explore this direction more deeply, and consider as well whether the features of the WPE energy landscape lend themselves to the demonstration of fundamental speedup using emerging quantum annealing devices, in particular those whose classical simulation is known to be intractable. It is our hope that the insights into the ensemble's physical and algorithmic classical properties presented in this work will solidly underpin these future directions.

## VI. ACKNOWLEDGEMENTS

We acknowledge useful discussions with Bill Macready, Federico Ricci-Tersenghi, Hidetoshi Nishimori, Mohammad Amin, Andrew King, Salvatore Mandrà, Cathy McGeogh, and Aidan Roy. H. G. K. would like to thank the Department of Poultry Science at Texas A&M University for providing emotional support. H. G. K. and C. P. are supported in part by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), via MIT Lincoln Laboratory Air Force Contract No. FA8721-05-C-0002. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of ODNI, IARPA, or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purpose notwithstanding any copyright annotation thereon.

## VII. APPENDIX

### A. Cavity calculation of the TAP equations

For notational convenience, we re-define the Hamiltonian of the WPE to include the constant terms corresponding to the diagonal elements of  $\tilde{\mathbf{J}}$ ; this will not affect the final result. Further, without loss of generality we assume that the ferromagnetic solution  $\mathbf{t} = (1, 1, \dots, 1)$  and its reversal are to be planted as the ground state. The Hamiltonian of the WPE is

$$H(\mathbf{s}) = -\frac{1}{2} \sum_{i,j=1}^N \tilde{J}_{ij} s_i s_j,$$

where

$$\tilde{J}_{ij} = -\frac{1}{N} \sum_{\mu=1}^M w_i^\mu w_j^\mu$$

and  $\mathbf{w}^\mu \sim \mathcal{N}(\mathbf{0}, \Sigma)$ . As before,  $M = \alpha N$  for  $\alpha > 0$  and  $N \rightarrow \infty$ . The elements of the covariance matrix  $\Sigma$  are:

$$\Sigma_{ij} = \begin{cases} 1 & \text{for } i = j \\ -\frac{1}{N-1} & \text{for } i \neq j \end{cases}.$$

This covariance structure implies that the Gaussian process generating  $\mathbf{w}$  is not only stationary but *exchangeable*. As discussed in Sec. III, this model bears many resemblances to the Hopfield network, but in addition to the negation of the interactions, we have the additional complication of the “anti-patterns”  $\{\mathbf{w}^\mu\}$  consisting of correlated elements. Nonetheless, by closely following the two-step cavity-based method presented by Shamir and Sompolinski [51] for the Hopfield model, taking special care to account for the correlations, we may derive the TAP equations for the WPE.

The cavity approach [48] derives the self-consistent relation for each local spin magnetization by first considering removal of the spin from the  $N + 1$ -spin system and defining a state distribution on the  $N$ -spin subsystem. Remarkably, the joint distribution of the field and spin in the original system can be expressed in terms of the field distribution resulting from the subsystem, allowing the spin and field statistics for the full system to be related to those of the subsystem. On their own, these exact relations do not give much insight because they are intractable to compute, but when the subsystem field distribution can be justified to be Gaussian, the

TAP equations for the magnetizations may be obtained. A substantial amount of the work is in deriving the correct parameters for the field distribution. The details for the WPE follow.

#### 1. Cavitating a spin

Consider an  $(N + 1) \times (N + 1)$  WPE matrix  $\tilde{\mathbf{J}}$  coupling spins  $\{0, \dots, N\}$  through  $M$  vectors  $\{\mathbf{w}^\mu\}$ , i.e.  $\tilde{J}_{ij} = -\frac{1}{N+1} \sum_{\mu=1}^M w_i^\mu w_j^\mu$ . We can decompose this Hamiltonian into a sub-Hamiltonian  $H^{(N)}$  consisting only of interactions among spins  $\{1, \dots, N\}$ , denoted here by  $s_{1:N}$ , and a term accounting for the interaction between spin zero and the others:

$$H^{(N+1)}(\mathbf{s}) = H^{(N)}(s_{1:N}) - h_0(s_{1:N})s_0 + \frac{1}{2}\tilde{J}_{00}$$

with

$$H^{(N)}(s_{1:N}) = -\frac{1}{2} \sum_{i,j=1}^N \tilde{J}_{ij} s_i s_j$$

and

$$h_0(s_{1:N}) = \sum_{j=1}^N \tilde{J}_{0j} s_j.$$

The final constant in  $H^{(N+1)}$  is irrelevant and will be dropped. The exact joint distribution over  $(s_0, h_0)$  can be shown to be

$$P^{(N+1)}(h_0, s_0) = \frac{1}{\xi} \exp(\beta h_0 s_0) P^{(N)}(h_0),$$

where

$$\xi = \frac{Z_{N+1}}{Z_N} = \langle 2 \cosh \beta h_0 \rangle_N$$

and  $\langle \dots \rangle_N$  refers to thermal averaging with respect to  $P^{(N)}(s_{1:N}) = \frac{1}{Z_N} \exp(-\beta H^{(N)}(s_{1:N}))$ . From this, we obtain the (intractable) relations

$$\begin{aligned} \langle s_0 \rangle_{N+1} &= \sum_{s_0} \int s_0 P^{(N+1)}(h_0, s_0) dh_0 \\ &= \frac{\int [\exp(\beta h_0) - \exp(-\beta h_0)] P^{(N)}(h_0) dh_0}{\langle 2 \cosh \beta h_0 \rangle_N} \\ &= \frac{\langle \sinh(\beta h_0) \rangle_N}{\langle \cosh(\beta h_0) \rangle_N} \end{aligned} \quad (29)$$

and

$$\begin{aligned}
\langle h_0 \rangle_{N+1} &= \sum_{s_0} \int h_0 P^{(N+1)}(h_0, s_0) dh_0 \\
&= \frac{\int h_0 [\exp(\beta h_0) + \exp(-\beta h_0)] P^{(N)}(h_0) dh_0}{\langle 2 \cosh \beta h_0 \rangle_N} \\
&= \frac{\langle h_0 \cosh(\beta h_0) \rangle_N}{\langle \cosh(\beta h_0) \rangle_N}. \tag{30}
\end{aligned}$$

The next step is to compute the field statistics to be used following a Gaussian assumption for  $P^{(N)}(h_0)$ . As for the Hopfield model, define the field mean and variance as

$$\begin{aligned}
\langle h_0 \rangle_N &= \sum_{j=1}^N \tilde{J}_{0j} \langle s_j \rangle_N, \\
\langle (\delta h_0)^2 \rangle_N &= \langle h_0^2 \rangle_N - \langle h_0 \rangle_N^2.
\end{aligned}$$

Note that

$$\langle (\delta h_0)^2 \rangle_N = \sum_{i=1}^N \sum_{j=1}^N \tilde{J}_{0i} \tilde{J}_{0j} \chi_{ij}^{(N)},$$

where

$$\chi_{ij}^{(N)} \triangleq \langle \delta s_i \delta s_j \rangle = \langle s_i s_j \rangle_N - \langle s_i \rangle_N \langle s_j \rangle_N.$$

Using the definition of  $\tilde{J}_{0i}$ , we obtain

$$\langle (\delta h_0)^2 \rangle_N = \frac{1}{(N+1)^2} \sum_{\mu=1}^M \sum_{\nu=1}^M w_0^\mu w_0^\nu \sum_{i=1}^N \sum_{j=1}^N w_i^\mu w_j^\nu \chi_{ij}^{(N)}. \tag{31}$$

If we define the overlap of spins  $\{1, \dots, N\}$  with the last  $N$  components of  $\mathbf{w}^\mu$  as

$$\eta_\mu \triangleq \frac{1}{N} \sum_{i=1}^N w_i^\mu s_i$$

and its covariance under the cavitared spin distribution

$$\langle \delta \eta_\mu \delta \eta_\nu \rangle_N \triangleq \langle \eta_\mu \eta_\nu \rangle_N - \langle \eta_\mu \rangle_N \langle \eta_\nu \rangle_N,$$

we obtain

$$\langle \delta \eta_\mu \delta \eta_\nu \rangle_N = \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N w_i^\mu w_j^\nu \chi_{ij}^{(N)}$$

and hence

$$\langle (\delta h_0)^2 \rangle_N \approx \sum_{\mu=1}^M \sum_{\nu=1}^M w_0^\mu w_0^\nu \langle \delta \eta_\mu \delta \eta_\nu \rangle_N.$$

Noting  $\chi_{ii}^{(N)} = 1 - \langle s_i^2 \rangle_N = O(1)$  while  $\chi_{ij}^{(N)} = O(\frac{1}{\sqrt{N}})$  for  $i \neq j$ , we proceed to determine the magnitude of  $\langle \delta \eta_\mu \delta \eta_\nu \rangle_N$  in order to simplify the field variance, bearing in mind that while  $\mathbf{w}^\mu$  is independent of  $\mathbf{w}^\nu$ , there are componentwise correlations within each vector not present in the Hopfield model. The conclusion will be that just as for the Hopfield model,  $\langle \delta \eta_\mu \delta \eta_\nu \rangle_N$  is  $O(\frac{1}{N^{3/2}})$  when  $\mu \neq \nu$  and  $O(\frac{1}{N})$  and when  $\mu = \nu$ . Define

$$S_{\mu\nu} = \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N w_i^\mu w_j^\nu \chi_{ij},$$

where the superscript on  $\chi$  has been dropped. We seek  $\mathbb{E}[S_{\mu\nu}]$  and the fluctuations  $\sqrt{\mathbb{E}[S_{\mu\nu}^2] - \mathbb{E}[S_{\mu\nu}]^2}$ , where the expectations are taken over  $\{\mathbf{w}\}$ .

The linear expectations are straightforward to compute; when  $\mu \neq \nu$ , we have

$$\mathbb{E}[S_{\mu\nu}] = \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N \mathbb{E}[w_i^\mu] \mathbb{E}[w_j^\nu] \chi_{ij} = 0$$

while, recalling the covariance structure of  $\mathbf{w}$ , when  $\mu = \nu$ ,

$$\begin{aligned}
\mathbb{E}[S_{\mu\mu}] &= \frac{1}{N^2} \left[ \sum_{i=1}^N \mathbb{E}[(w_i^\mu)^2 \chi_{ii}] + \sum_{i \neq j} \mathbb{E}[w_i^\mu w_j^\mu \chi_{ij}] \right] \\
&= \frac{1}{N^2} \left[ O(N) - O\left(\frac{N(N-1)}{N\sqrt{N}}\right) \right] \\
&= O\left(\frac{1}{N}\right),
\end{aligned}$$

which is different from the Hopfield model, in which this quantity is zero.

Direct computation of the quadratic expectation  $\mathbb{E}[S_{\mu\nu}^2]$  is tedious but straightforward. In the expansion, there will be a total of  $N^4$  terms of the form  $\chi_{ij} \chi_{kl} w_i^\mu w_j^\nu w_k^\mu w_l^\nu$ . We first make a relevant partitioning of these terms for generic  $\{\mu, \nu\}$  and then compute the expectations for the cases where they are equal and different.

- **Case**  $i = k, j = l$

If  $i = j$ , there are  $N$  terms of the form  $\chi_{ii}^2(w_i^\mu)^2(w_i^\nu)^2$

If  $i \neq j$ , there are  $N(N-1)$  terms like  $\chi_{ij}^2(w_i^\mu)^2(w_j^\nu)^2$

- **Case**  $i = k, j \neq l$

If  $i = j$ , there are  $N(N-1)$  terms as  $\chi_{ii}\chi_{il}(w_i^\mu)^2w_i^\nu w_l^\nu$

If  $i \neq j$  we have  $N$  terms like  $\chi_{ij}\chi_{ii}(w_i^\mu)^2w_i^\nu w_j^\nu$  and  $N(N-2)$  terms as  $\chi_{ij}\chi_{il}(w_i^\mu)^2w_j^\nu w_l^\nu$

- **Case**  $i \neq k, j = l$

This is a rotated version of the previous case. When  $i = j$ , there are  $N(N-1)$  terms of the form  $\chi_{jj}\chi_{kj}w_i^\mu w_k^\mu(w_j^\nu)^2$

When  $i \neq j$ , there are  $N$  terms like  $\chi_{ij}\chi_{jj}w_i^\mu w_j^\mu(w_j^\nu)^2$  and  $N(N-2)$  terms as  $\chi_{ij}\chi_{kj}w_i^\mu w_k^\mu(w_j^\nu)^2$

- **Case**  $i \neq k, j \neq l$

When  $i = j$ , there are  $N(N-1)$  terms like  $\chi_{ii}\chi_{kk}w_i^\mu w_k^\mu w_i^\nu w_k^\nu$  and  $N(N-1)(N-2)$  terms as  $\chi_{ii}\chi_{kl}w_i^\mu w_k^\mu w_i^\nu w_l^\nu$

When  $i \neq j$ , there are  $N(N-1)(N-2)$  terms as  $\chi_{ij}\chi_{kk}w_i^\mu w_k^\mu w_j^\nu w_l^\nu$  and  $N(N-1)(N^2-3N+3)$  terms like  $\chi_{ij}\chi_{kl}w_i^\mu w_k^\mu w_j^\nu w_l^\nu$

Consider now the case of  $\mu \neq \nu$ . Adding the terms in the expansion, recalling that  $\mathbf{w}^\mu$  is independent of  $\mathbf{w}^\nu$  and that  $\mathbb{E}[w_i w_j] = -\frac{1}{N}$  when  $i \neq j$ , we find that

$$\mathbb{E}[S_{\mu\nu}^2] = O\left(\frac{1}{N^3}\right),$$

implying that  $S_{\mu\nu}$  is  $O\left(\frac{1}{N^{3/2}}\right)$ . This quantity is of identical order in the Hopfield model despite the correlations in  $\{\mathbf{w}\}$ .

When  $\mu = \nu$  on the other hand, we find that

$$\begin{aligned} \mathbb{E}[S_{\mu\mu}] &= \frac{1}{N^4} \left[ NO(1)\mathbb{E}[w_i^4] + N(N-1)O\left(\frac{1}{N}\right) \dots \right. \\ &\quad \mathbb{E}[w_i^2 w_j^2] + 2N(N-1)O\left(\frac{1}{\sqrt{N}}\right)\mathbb{E}[w_i^3 w_l] + 2N \dots \\ &\quad O\left(\frac{1}{\sqrt{N}}\right)\mathbb{E}[w_i^3 w_j] + 2N(N-2)O\left(\frac{1}{N}\right)\mathbb{E}[w_i^2 w_j w_l] + \dots \\ &\quad N(N-1)O(1)\mathbb{E}[w_i^2 w_k^2] + N(N-1)(N-2) \dots \\ &\quad O\left(\frac{1}{\sqrt{N}}\right)\mathbb{E}[w_i^2 w_k w_l] + N(N-1)(N-2)O\left(\frac{1}{\sqrt{N}}\right) \dots \\ &\quad \mathbb{E}[w_i w_j w_k^2] + N(N-1)(N^2-3N+3)O\left(\frac{1}{\sqrt{N}}\right) \dots \\ &\quad \left. \mathbb{E}[w_i w_j w_k w_l] \right], \end{aligned} \quad (32)$$

where the  $O(\dots)$  terms in Eq. (32) refer to the effects of the  $\{\chi_{ij}\}$  products. Using the properties of the higher-order moments of a correlated Gaussian distribution (e.g., Ref. [68]), we have by the exchangeability of our distribution for any indices such that different letters refer to different values,

$$\begin{aligned} \mathbb{E}[w_i^4] &= 3 = O(1) \\ \mathbb{E}[w_i^3 w_j] &= -\frac{3}{N} = O\left(\frac{1}{N}\right) \\ \mathbb{E}[w_i^2 w_j w_k] &= -\frac{1}{N} + \frac{2}{N^2} = O\left(\frac{1}{N}\right) \\ \mathbb{E}[w_i^2 w_j^2] &= 1 + \frac{2}{N^2} = O(1) \\ \mathbb{E}[w_i w_j w_k w_l] &= -\frac{3}{N^2} = O\left(\frac{1}{N^2}\right). \end{aligned}$$

Substituting these into Eq. (32), we obtain the result that  $\mathbb{E}[S_{\mu\mu}^2] = O(1/N^2)$  and hence that  $S_{\mu\mu}$  is typically

$$\sqrt{O\left(\frac{1}{N^2}\right) - \frac{1}{N^2}} \pm \frac{1}{N}, \quad (33)$$

namely  $O(1/N)$ . These results imply that the field variance in Eq. (31)

$$\langle (\delta h_0)^2 \rangle_N = \sum_{\mu=1}^M (w_0^\mu)^2 \langle (\delta \eta_\mu)^2 \rangle_N + \sum_{\mu \neq \nu} w_0^\mu w_0^\nu \langle \delta \eta_\mu \delta \eta_\nu \rangle_N$$

is

$$O\left(\frac{M}{N}\right) + O\left(\frac{\sqrt{M(M-1)}}{N^{3/2}}\right) = O(1) + O\left(\frac{1}{\sqrt{N}}\right),$$

where the order of the second summation follows from the independence of  $\mu$  and  $\nu$ ; by self-averaging it can hence be approximated by

$$\langle(\delta h_0)^2\rangle_N = \sum_{\mu=1}^M \langle(\delta \eta_\mu)^2\rangle_N \triangleq V_N. \quad (34)$$

Finally, recall that  $h_0 = -\frac{1}{N+1} \sum_{\mu=1}^M w_0^\mu \sum_{i=1}^N w_i^\mu s_i \approx -\sum_{\mu=1}^M w_0^\mu \eta_\mu$ . The  $\{\eta_\mu\}$  decorrelate at the same rate as they do in the Hopfield model; further,  $w_0^\mu$  follows deterministically from  $\{w_{1:N}^\mu\}$  and adds no information about the state distribution on sites  $\{1, \dots, N\}$  and hence on the  $\{\eta_\mu\}$ . This suggests that the field can be approximated by a limiting Gaussian distribution:

$$P^{(N)}(h_0) = \frac{1}{\sqrt{2\pi V}} \exp \left[ -\frac{(h_0 - \langle h_0 \rangle_N)^2}{2V} \right],$$

where  $V \triangleq \lim_{N \rightarrow \infty} V_N$ . This Gaussian approximation dramatically simplifies relations in Eqs. (29) and (30), which reduce to

$$\begin{aligned} \langle s_0 \rangle_{N+1} &= \tanh(\beta \langle h_0 \rangle_N) \\ \langle h_0 \rangle_{N+1} &= \langle h_0 \rangle_N + \beta V \langle s_0 \rangle_{N+1}. \end{aligned} \quad (35)$$

Considering deletion of any spin  $i$  rather than zero, the TAP relation

$$\langle s_i \rangle = \tanh \left[ \beta \left( \sum_{j \neq i} J_{ij} \langle s_j \rangle - \beta V \langle s_i \rangle \right) \right] \quad (36)$$

follows. Note that the distinction between  $\mathbf{J}$  and  $\tilde{\mathbf{J}}$  disappears at this point; a consequence of the cavity method is that  $\{\tilde{J}_{ii}\}$  terms are disregarded as one expects. To fully specify the TAP relation,  $V$  must be determined for the WPE; we turn to this task next.

## 2. Cavitating a $\mathbf{w}$

Consider now a system of  $N$  spins but with  $M+1$   $\{\mathbf{w}^\mu\}$ . The corresponding Hamiltonian can be related to that of a system with a single  $\mathbf{w}^\mu$  removed. Specifically, if

$$H^{(M)}(\mathbf{s}) = \frac{1}{2N} \sum_{i,j=1}^N \sum_{\mu=1}^M w_i^\mu w_j^\mu s_i s_j$$

then the full Hamiltonian is

$$\begin{aligned} H^{(M+1)}(\mathbf{s}) &= H^{(M)}(\mathbf{s}) + \frac{1}{2N} \left( \sum_{i=1}^N w_i^0 s_i \right) \left( \sum_{j=1}^N w_j^0 s_j \right) \\ &= H^{(M)}(\mathbf{s}) + \frac{N}{2} (\eta_0)^2, \end{aligned}$$

where again an irrelevant constant has been dropped. Now following the cavity procedure, we obtain the distribution of  $\eta_0$  relative to the Boltzmann distribution corresponding to  $H^{(M+1)}$  in terms of that corresponding to  $H^{(M)}$ :

$$\begin{aligned} P^{(M+1)}(\mathbf{s}) &= \frac{1}{Z_{M+1}} \sum_{\mathbf{s}} \exp \left[ -\beta H^{(M+1)}(\mathbf{s}) \right] \delta \left( \eta_0 - \frac{1}{N} \sum_{i=1}^N w_i^0 s_i \right) = \\ &= \frac{1}{Z_{M+1}} \sum_{\mathbf{s}} \exp \left[ -\beta H^{(M)}(\mathbf{s}) + \frac{N}{2} (\eta_0)^2 \right] \delta \left( \eta_0 - \frac{1}{N} \sum_{i=1}^N w_i^0 s_i \right) = \\ &= \frac{Z_M}{Z_{M+1}} \exp \left[ -\beta \frac{N}{2} (\eta_0)^2 \right] \sum_{\mathbf{s}} \left[ P^{(M)}(\mathbf{s}) \delta \left( \eta_0 - \frac{1}{N} \sum_{i=1}^N w_i^0 s_i \right) \right] = \\ &= \frac{Z_M}{Z_{M+1}} \exp \left[ -\beta \frac{N}{2} (\eta_0)^2 \right] P^{(M)}(\eta_0). \end{aligned} \quad (37)$$

The expectation  $\langle \eta_0 \rangle_M = \frac{1}{N} \sum_{i=1}^N w_i^0 \langle s_i \rangle_M = 0$  by self-averaging due to the stationarity of  $\mathbf{w}^0$ . The variance  $\langle (\delta \eta_0)^2 \rangle_M$  was shown to be  $O\left(\frac{1}{N}\right)$  previously; for large  $N$ , its typical value is

$$\begin{aligned} \langle (\eta_0)^2 \rangle_M &= \frac{1}{N^2} \left[ \sum_{i \neq j} w_i^0 w_j^0 \chi_{ij}^{(M)} + \sum_{i=1}^N \chi_{ii}^{(M)} \right] \\ &\rightarrow \frac{1}{N^2} \left[ -\sum_{i \neq j} \frac{1}{N-1} \chi_{ij}^{(M)} + \sum_{i=1}^N (1 - \langle s_i \rangle_M^2) \right] \\ &\rightarrow \frac{1}{N} (1 - q), \end{aligned}$$

where

$$q = \frac{1}{N} \sum_{i=1}^N \langle s_i \rangle_M^2.$$



Finally, assuming that  $P^{(M)}(\eta_0)$  is Gaussian and using the relation in Eq. (37), we obtain

$$\begin{aligned} P^{(M+1)}(\eta_0) &\propto \exp \left[ -\frac{N}{2(1-q)}\eta_0^2 - \frac{N\beta}{2}\eta_0^2 \right] \\ &= \exp \left[ -\frac{1}{2} \left[ \frac{N}{1-q} + N\beta \right] \eta_0^2 \right] \end{aligned}$$

which is a Gaussian with mean zero and variance

$$\langle (\delta\eta_0)^2 \rangle_{M+1} = \frac{1-q}{N(1+\beta(1-q))}.$$

Note that this value would result for any  $\mathbf{w}^\mu$  removed. To obtain the field variance, we use Eq. (34) to obtain

$$\begin{aligned} V &= \sum_{\mu=1}^M \langle (\delta\eta_\mu)^2 \rangle_N \\ &= \alpha N \langle (\delta\eta_0)^2 \rangle_{M+1} \\ &= \frac{\alpha(1-q)}{1+\beta(1-q)}. \end{aligned}$$

This resembles that of the Hopfield model but with a changed sign in the denominator.

### B. Limiting spectral distribution of $\mathbf{J}$

We determine the limiting eigenvalue distribution of the  $\mathbf{J}$  matrix. To do so, it suffices to determine the spectral distribution of  $\tilde{\mathbf{J}}$ . To see why, first note that  $\tilde{J}_{ii} \rightarrow -\alpha$  for all  $i$  by the law of large numbers. Recalling that  $\mathbf{J} = \tilde{\mathbf{J}} - \text{diag}(\tilde{\mathbf{J}})$ , this implies that  $\mathbf{J}$  is asymptotically related to  $\tilde{\mathbf{J}}$  by addition of a uniform quantity to the diagonal, namely

$$\mathbf{J} = \tilde{\mathbf{J}} + \alpha \mathbf{I}_N,$$

which in turn means that the eigenvalues of  $\mathbf{J}$  are simply those of  $\tilde{\mathbf{J}}$  translated by  $\alpha$ .

For Wishart matrices of the form  $\frac{1}{M}\mathbf{X}\mathbf{X}^T$ , where  $\mathbf{X}$  are  $N \times M$  matrices whose elements are independent zero-mean unit-variance Gaussian variates and  $M = \alpha N$ , the limiting spectral distribution is known as the Marchenko-Pastur [53] law. The fact that the columns of  $\mathbf{W}$  are correlated Gaussian variates seems at first to complicate the determination for  $\frac{1}{N}\mathbf{W}\mathbf{W}^T$ . The structure of the specific covariance matrix  $\Sigma$  considerably simplifies matters,

however. Recalling that

$$\Sigma = \frac{N}{N-1} [\mathbf{I} - \mathbf{t}\mathbf{t}^T],$$

it is apparent that first,  $\mathbf{t}$  is an eigenvector with null eigenvalue, and second, that any vector in the subspace orthogonal to  $\mathbf{t}$  is an eigenvector with eigenvalue  $N/(N-1) \rightarrow 1$ . Hence, any orthonormal set of  $N-1$  vectors orthogonal to  $\mathbf{t}$  can be used to represent  $\Sigma$ , and the variation of  $\mathbf{w}$  along each of these eigenvectors is asymptotically of unit magnitude. The procedure for generating  $\mathbf{w}$  is for large  $N$  thus equivalent to first generating vector  $\mathbf{x}$  whose first  $N-1$  elements are independently  $\sim \mathcal{N}(0, 1)$  and whose last element is zero and next, transforming  $\mathbf{x}$  by some unitary matrix  $\mathbf{U}$  rotating the  $N^{\text{th}}$  coordinate vector  $\mathbf{e}_N = (0, \dots, 0, 1)$  to  $\mathbf{t}$ . This implies that the spectral distribution of  $\mathbf{W}\mathbf{W}^T$  approaches that of  $\mathbf{X}\mathbf{X}^T$ , where

$$\mathbf{X} = \begin{bmatrix} \tilde{\mathbf{X}} \\ \mathbf{0}^T \end{bmatrix}$$

consists of an  $(N-1) \times M$  matrix  $\tilde{\mathbf{X}}$  composed of iid normal variates and a final row of zeros; the eigenvalues of  $\mathbf{X}\mathbf{X}^T$  are thus those of  $\tilde{\mathbf{X}}\tilde{\mathbf{X}}^T$  with an extra zero added. The limiting spectral distribution of  $\frac{1}{N-1}\tilde{\mathbf{X}}\tilde{\mathbf{X}}^T$  can be straightforwardly obtained from the Marchenko-Pastur law by appropriate change of variables. We then obtain the limiting eigenvalue distribution of the matrix  $-\tilde{\mathbf{J}} = \frac{1}{N}\mathbf{W}\mathbf{W}^T$  to be

$$\tilde{f}(\lambda) = \begin{cases} (1-\alpha)\delta(\lambda) + \tilde{f}_+(\lambda) & \alpha < 1 \\ \frac{1}{N}\delta(\lambda) + \frac{N-1}{N}\tilde{f}_+(\lambda) & \alpha \geq 1 \end{cases}, \quad (38)$$

where

$$\tilde{f}_+(\lambda) = \frac{1}{2\pi} \frac{\sqrt{(\lambda_+ - \lambda)(\lambda - \lambda_-)}}{\lambda} \mathbf{1}[\lambda \in [\lambda_-, \lambda_+]]$$

and

$$\begin{aligned} \lambda_- &= \alpha - 2\sqrt{\alpha} + 1 \\ \lambda_+ &= \alpha + 2\sqrt{\alpha} + 1. \end{aligned}$$

Note that the delta spike at zero never disappears, a feature that turns out to crucially influence the phase behavior for large  $\alpha$ . Finally, the spectral distribution of  $\mathbf{J} = \tilde{\mathbf{J}} + \alpha \mathbf{I}$  follows by reflection and translation:

$$f(\lambda) = \begin{cases} (1-\alpha)\delta(\lambda - \alpha) + f_+(\lambda) & \alpha < 1 \\ \frac{1}{N}\delta(\lambda - \alpha) + \frac{N-1}{N}f_+(\lambda) & \alpha \geq 1 \end{cases}, \quad (39)$$

where

$$f_+(\lambda) = -\frac{1}{2\pi} \frac{\sqrt{(\lambda_+ - \lambda)(\lambda - \lambda_-)}}{\lambda - \alpha} \mathbf{1}[\lambda \in [\lambda_-, \lambda_+]]$$

and

$$\begin{aligned}\lambda_- &= -2\sqrt{\alpha} - 1 \\ \lambda_+ &= 2\sqrt{\alpha} - 1.\end{aligned}$$

The distribution consists of a continuous component over the support  $[\lambda_-, \lambda_+]$  and a persistent delta function at  $\alpha$ , where of course  $\alpha \geq \lambda_+$ .

### C. Energy histogram of the Wishart ensemble

Define the  $M$ -vector of normalized state overlaps with the  $\{\mathbf{w}^\mu\}$

$$\boldsymbol{\eta}(\mathbf{s}) = \frac{1}{\sqrt{2N}} \mathbf{W}^T \mathbf{s}$$

so that  $H(\mathbf{s}) = \boldsymbol{\eta}^T \boldsymbol{\eta}$  and

$$\begin{aligned}p_E(e|\{\mathbf{w}^\mu\}) &= \Pr(\boldsymbol{\eta}^T \boldsymbol{\eta} = e|\{\mathbf{w}^\mu\}) \\ &= \frac{1}{2^N} \sum_{\mathbf{s}} \delta\left(e - \frac{1}{2N} \mathbf{s}^T \mathbf{W} \mathbf{W}^T \mathbf{s}\right).\end{aligned}$$

The derivation of the marginal energy

$$p_E(e) \triangleq \int_{\{\mathbf{w}^\mu\}} p_E(e|\{\mathbf{w}^\mu\}) f(\mathbf{w}^1, \dots, \mathbf{w}^M) d\{\mathbf{w}^\mu\}$$

is simplified by decomposing the integration into sums of expectations over subsets of states with a constant number of positive elements and exploiting the exchangeability of  $f(\mathbf{w})$ . Let  $N_+(\mathbf{s})$  be the number of elements in  $\mathbf{s}$  with value  $+1$ ; for a magnetization  $m \in [-1, 1]$ ,  $N_+ = \frac{(1+m)N}{2}$ . On the  $N_+$  constrained subset, the marginal density of  $E$  is

$$\begin{aligned}p_E(e|N_+) &= \frac{1}{\binom{N}{N_+}} \mathbb{E}_{\{\mathbf{w}^i\}} \left[ \sum_{\mathbf{s}: N_+} \delta\left(e - \frac{1}{2N} \mathbf{s}^T \mathbf{W} \mathbf{W}^T \mathbf{s}\right) \right] \\ &= \frac{1}{\binom{N}{N_+}} \sum_{\mathbf{s}: N_+} \mathbb{E}_{\{\mathbf{w}^i\}} \left[ \delta\left(e - \frac{1}{2N} \mathbf{s}^T \mathbf{W} \mathbf{W}^T \mathbf{s}\right) \right],\end{aligned}$$

where the sums are over states with  $N_+$  positive entries. The joint distribution

$$f(\mathbf{w}^1, \dots, \mathbf{w}^M) = f(\mathbf{w}^1), \dots, f(\mathbf{w}^M)$$

by the independence of the columns, but the components of each  $\mathbf{w}^\mu$  are correlated. They are however exchangeable, and since the same  $\mathbf{s}$  appears in each term in the sum this implies that

$$\mathbb{E}_{\{\mathbf{w}^i\}} \left[ \delta\left(e - \frac{1}{2N} \mathbf{s}^T \mathbf{W} \mathbf{W}^T \mathbf{s}\right) \right] \quad (40)$$

only depends on  $N_+$ . Consider the specific case of  $\mathbf{s}$  whose first  $N_+$  elements are 1. Define the sums (one for each column of  $\mathbf{W}$ )

$$A_+^\mu \triangleq \frac{1}{\sqrt{2N}} \sum_{j=1}^{N_+} w_j^\mu.$$

Because  $\sum_{i=1}^N w_i^\mu = 0$ , the sum

$$\begin{aligned}A_-^\mu &\triangleq \frac{1}{\sqrt{2N}} \sum_{j=N_++1}^N w_j^\mu \\ &= -A_+^\mu\end{aligned}$$

deterministically. By the independence of the  $\mathbf{w}^\mu$  the sums  $\{A_+^\mu\}$  are independent random variables and from the properties of linear transformations of Gaussian variables each is distributed according to a zero-mean Gaussian with variance

$$\sigma_{N_+}^2 = \frac{N_+(N - N_+)}{2N(N - 1)}.$$

When  $N_+ = 0$  or  $N_+ = N$ , the zero-variance Gaussian is defined to be a delta function as we expect. We express the expectation (40) at fixed  $N_+$  as

$$\begin{aligned}\mathbb{E}_{\{\mathbf{w}^i\}} &\left[ \delta\left(e - \frac{1}{2N} \mathbf{s}^T \mathbf{W} \mathbf{W}^T \mathbf{s}\right) \right] \\ &= \mathbb{E}_{\{A_+^\mu, A_-^\mu\}} \left[ \delta\left(e - \sum_{\mu=1}^M (A_+^\mu - A_-^\mu)^2\right) \right] \\ &= \mathbb{E}_{\{A_+^\mu\}} \left[ \delta\left(e - \sum_{\mu=1}^M (2A_+^\mu)^2\right) \right].\end{aligned} \quad (41)$$

Now let

$$S \triangleq \sum_{\mu=1}^M (A_+^\mu)^2$$

with density  $f_S$ . We then have

$$\begin{aligned} & \mathbb{E}_{\{A_+^i\}} \left[ \delta \left( e - \sum_{i=1}^M (2A_+^i)^2 \right) \right] \\ &= \mathbb{E}_S \left[ \delta \left( e - 4S \right) \right] \\ &= \frac{1}{4} f_S \left( \frac{e}{4} \right). \end{aligned} \quad (42)$$

Because  $S$  is the sum of squares of  $M$  zero-mean iid Gaussians with variance  $\sigma_{N_+}^2$ , it is gamma-distributed, i.e., with density

$$f_S(s) = \begin{cases} \frac{1}{\Gamma(k)\theta^k} s^{k-1} \exp\left(-\frac{s}{\theta}\right) & \text{for } s \geq 0 \\ 0 & \text{for } s < 0 \end{cases}$$

with parameters  $(k, \theta) = (M/2, 2\sigma_{N_+}^2)$ . For large  $N$  and  $m$  relating  $N_+$  to  $N$ ,  $\sigma_{N_+}^2 \approx \frac{1-m^2}{8}$ . We then obtain the constrained- $m$  energy histogram

$$\begin{aligned} p_E(e|m) &= \frac{1}{4} f_S \left( \frac{e}{4} \right) \\ &= \frac{1}{\Gamma(M/2)(1-m^2)^{M/2}} e^{M/2-1} \exp\left(-\frac{e}{1-m^2}\right) \end{aligned}$$

for  $e \geq 0$  and zero otherwise and the overall energy distribution

$$p_E(e) = \frac{1}{2^N} \sum_m \left( \frac{N}{(1+m)N} \right) p_E(e|m), \quad (43)$$

where the sum runs over the  $N+1$  values of  $m$  mapping to  $N_+ \in \{0, \dots, N\}$ . As described, for example, in Refs. [42, 54], applying Stirling's large- $N$  approximation to the binomial coefficients, replacing the sum with an integral and evaluating it with Laplace's method, we obtain

$$p_E(e) \approx p_E(e|m=0)$$

or the marginal gamma energy density for the WPE

$$p_E(e) = \begin{cases} \frac{1}{\Gamma(M/2)} e^{M/2-1} \exp(-e) & \text{for } e \geq 0 \\ 0 & \text{for } e < 0. \end{cases} \quad (44)$$

#### D. Annealed approximation

The purpose of the annealed approximation is to provide a simple bound on the typical case behaviour of the free

energy at leading order in  $N$ . The free energy density is defined

$$f = \lim_{N \rightarrow \infty} -\frac{1}{\beta N} \left\langle \log \sum_{\mathbf{s}} \exp[-\beta H(\mathbf{s})] \right\rangle, \quad (45)$$

where  $\langle \cdot \rangle$  denotes an average over instances. Discontinuities in the free energy describe the phase transitions, and derivatives describe the order parameter(s) and other statistically significant quantities at thermal equilibrium.

It is convenient for this section to consider a model defined by the Hamiltonian

$$H(\mathbf{s}) = \frac{1}{N} \sum_{\mu=1}^{\alpha N} \left( \sum_{i',i} Z_{i',\mu} \left[ \delta_{i,i'} - \frac{\kappa}{N} \right] s_i \right)^2, \quad (46)$$

where  $Z_{i,\mu}$  are independent and normally distributed random variables. For the case  $\kappa = 1$  this Hamiltonian is identical to the main text Hamiltonian up to the choice of the embedding solution ( $t_i = 1, \forall i$ ), inclusion of a nonzero diagonal term in the coupling matrix (which adds a constant offset to the energy), and corrections of  $O(1/N)$ . These restrictions are for the convenience of analysis, and have no significant impact on the analysis method or conclusions of the section.

The parameter  $\kappa$  is useful in making the connection to the anti-Hopfield model (obtained for the case  $\kappa = 0$ ) [36], and in identifying a variation on the principle of embedding discussed in the main text. Tuning of this parameter allows one to control the energy level of the planted solution at leading order in  $N$ , allowing an embedding of a ground state with control over the gap at fixed  $\alpha$ , or implanting an excited state well separated from other stable and metastable states.

The annealed approximation may be used to obtain a lower bound,  $f \geq f_A$ , on the free-energy density

$$f_A = \lim_{N \rightarrow \infty} -\frac{1}{\beta N} \log \left\langle \sum_{\mathbf{s}} \exp[-\beta H(\mathbf{s})] \right\rangle. \quad (47)$$

The physical interpretation for this approximation is that the quenched degrees of freedom ( $Z$ ) are treated on an equal footing with the dynamical degrees of freedom ( $\mathbf{s}$ ). This means that models of lower energy can be selected disproportionately, since spin and model variables can become correlated to lower the free energy. By this process it is possible that atypical models can dominate the free energy so that typical case is not reflected.

However, in this ensemble we show that the free energy is correct through much of the phase space in agreement with the TAP analysis of Sec. III A.

Owing to the quadratic form of the Hamiltonian in the order parameter, it is rather straightforward to take the disorder average explicitly, which yields:

$$f_A = \lim_{N \rightarrow \infty} -\frac{1}{\beta N} \log \sum_{\mathbf{s}} \exp \{ \alpha N \text{Tr} \log [\mathbf{I} + \beta \mathbf{X}(\mathbf{s})] \}, \quad (48)$$

where  $X_{ij}(s) = \frac{1}{N}(s_i - \kappa \sum_{i'} s_{i'})(s_j - \kappa \sum_{j'} s_{j'})$ . Following this, we notice that the eigenvalues of  $\mathbf{X}$  are a function only of the sum of spin variables, thus the trace log can be simplified, defining  $m = \frac{1}{N} \sum s_i$  we can write

$$f_A = \lim_{N \rightarrow \infty} -\frac{1}{\beta N} \log \sum_m \binom{N}{(1+m)N/2} \exp \{ \alpha N \text{Tr} \log [1 + \beta(1 - m^2 \kappa(2 - \kappa))] \}. \quad (49)$$

Using Stirlings's approximation, and using a continuum approximation for  $m$  we can write

$$f_A = \lim_{N \rightarrow \infty} -\frac{1}{\beta N} \log \int dm \exp \left\{ \frac{N}{2} \log \left( \frac{1+m}{1-m} \right) + \alpha N \text{Tr} \log [1 + \beta(1 - m^2 \kappa(2 - \kappa))] \right\} \quad (50)$$

Finally, applying Laplace's approximation yields the result

$$f \geq \arg \max_x \left[ \frac{1}{2\beta} \log \frac{1+m}{1-m} + \dots \right. \\ \left. \alpha \log [1 + \beta(1 - m^2 \kappa(2 - \kappa))] \right]. \quad (51)$$

The quantity  $m$  can be readily associated with the planted state overlap (magnetization, for the case of  $t_i = 1 \forall i$ ). The first term is the standard mean-field entropy term, the latter term being an energy term.

Maximizing this equation involves solving for  $df/dm = 0$ , also called the saddle-point equation. At high temperature there is only the solution  $m = 0$ , whereas at low temperature (for large enough  $\kappa$ ) there are two additional solutions. For the case  $\kappa = 1$  of the main text [Eq. (51)] is identical to the dominant minima of the TAP equation free energy [Eq. (14)].

At low temperature it is interesting to observe that the crystalization transition occurs due to a competition between the energy term and the entropy term, one dominating in each regime. For the paramagnetic solution

is defined by  $m = 0$ , whereas for the planted solution  $m \approx 1$ . Equating these two terms for the case  $\kappa = 1$  we can find a simple approximation to the first order transition, the approximation in Eq. (20) proves a very accurate lower bound for the critical temperature shown in Fig. 1 for small values of  $\alpha$ .

In this section we have found that the annealed approximation recovers the more general TAP result derived in Sec. III A. For the special case of  $O(N)$  planting  $\kappa < 1$  we find that the embedded state remains thermodynamically dominant at low enough temperature for large enough  $\kappa$ , at a diminished critical temperature (and increased ground state energy). For the case  $\kappa \approx 0$ , and at small  $\alpha$  the annealed approximation is insufficient to predict all features of the phase diagram, for this we can use the replica method of Sec. VII E.

### E. The replica method

The annealed approximation of Sec. VIID, and TAP analysis presented (Sec. III A), are insufficient to describe all the features of our model phase space. To go beyond these we here introduce a replica method. The replica method when solved is able to capture non trivial properties of the exact free energy, at leading order in  $N$ , in a variety of disordered models closely related to our proposed model [25, 36, 52]. As presented here, the replica method is a nonrigorous method for purposes of obtaining insight through related models.

The replica method for the anti-Hopfield model ( $\kappa = 0$ ) is demonstrated by Nokura *et al.* [36], and may be derived along similar lines as the Hopfield model [25]. Only minor variations to the form of the Hopfield replica method (free energy) are necessary to analyze the WPE, but these changes have significant consequences.

The free energy density [Eq. (45)] can be rewritten for purposes of the replica method in terms of a replicated partition function

$$f = \lim_{N \rightarrow \infty} -\frac{1}{\beta N} \lim_{n \rightarrow 0} \frac{\partial}{\partial n} \langle \mathcal{Z}^n \rangle, \quad (52)$$

where  $\mathcal{Z}$  is the partition function, and the trick is to solve for the case of positive integer  $n$  and analytically continue to real  $n$ . For integer  $n$  we can write

$$\mathcal{Z}^n = \sum_{\mathbf{s}} \prod_{\rho=1}^n \exp(H(\mathbf{s}^\rho)) \quad (53)$$

using Eq. (46),  $\mathbf{s}^\rho$  being a vector of dimension  $N$ . From here we can follow closely the method in Section 2 of Ref. [25], which is considered a standard approach. This being understood, we are sparse in our derivation. The main difference in the method is the introduction of an additional order parameter  $x$  through the identity

$$\begin{aligned} & \int dm^\rho \delta(m^\rho - \frac{1}{N} \sum_i s_i^\rho) \\ &= \int dm^\rho d\hat{m}^\rho \exp \left( \hat{m}^\rho [Nm^\rho - \sum_i s_i^\rho] \right). \end{aligned} \quad (54)$$

The integral identity, and scaling with  $N$ , being appropriate for the large system limit saddle-point, to be later identified. A similar identity is introduced for the overlap  $q^{\rho,\rho'} = \frac{1}{N} \sum_i s_i^\rho s_i^{\rho'}$ . The replicated partition function following these manipulations be written

$$\begin{aligned} \langle Z^n \rangle &= \exp \left\{ \frac{\alpha N}{2} \text{Tr} \log [(1-\beta)\mathbf{I} + \beta\mathbf{q} - \dots \right. \\ &\quad \left. \beta\kappa(2-\kappa)\mathbf{m}\mathbf{m}^T] \right\} - \frac{N}{2} \sum_{\rho \neq \rho'} q_{\rho,\rho'} \hat{q}_{\rho,\rho'} - \dots \\ &\quad N \sum_\rho m_\rho \hat{m}_\rho + \log \text{Tr}_s \exp \left( \frac{1}{2} \mathbf{s}^T \hat{\mathbf{q}} \mathbf{s} + \hat{\mathbf{m}}^T \mathbf{s} \right) \Big\}, \end{aligned} \quad (55)$$

where  $\mathbf{I}$  is the  $n \times n$  identity matrix,  $\mathbf{q}$  and  $\hat{\mathbf{q}}$  are matrices of the same dimension with zero on diagonal (overlap order parameters), and  $\mathbf{m}$ ,  $\hat{\mathbf{m}}$  and  $\mathbf{s}$  are  $n \times 1$  vectors (alignment with planted solution order parameters).

For the special case  $m^\rho = 0$  (or  $\kappa = 0$ ) this free energy is identical to that of the anti-Hopfield model [36]. The interpretation of this result is as follows: since  $m$  describes the degree of alignment with the planted solution, we can argue that when there is no extensive alignment with the planted solution, such as at high temperature (above the first order planting transition) or in general for the space orthogonal to the planted solution, the model phenomena will be identical to the anti-Hopfield model. This is interesting, since in the anti-Hopfield model, and in related classes of problems, there is understood to be a replica symmetry breaking phenomena [52]. The space becomes divided into modes separated by extensive barriers at sufficiently low temperature. We thus expect this same phenomena to carry over to our model either as a stable or metastable solution. To determine which we must analyse the free energy within an approximation.

In this Appendix we will develop only the replica symmetric theory, this being understood as sufficient for the paramagnetic phase, and the crystal phase discussed in the main text.

In the *replica symmetric* solution we take  $m^\rho = m, \forall \rho$  and that  $q^{\rho,\rho'} = q$  for  $\rho \neq \rho'$  (the diagonal term is 0). The free energy at leading order can be determined from the saddle-point free energy

$$\begin{aligned} \beta f(q, \hat{q}, m, \hat{m}) &= -\frac{\alpha}{2} \left\{ \log(1 + \beta(1-q)) + \dots \right. \\ &\quad \left. + \frac{\beta(q - (2\kappa - \kappa^2)m^2)}{1 + \beta(1-q)} \right\} + \hat{m}m - \frac{1}{2}\hat{q}q + \frac{1}{2}\hat{q} - \dots \\ &\quad \int dz \exp \left( -\frac{z^2}{2} \right) \log \left[ 2 \cosh \left( \sqrt{\hat{q}}z + \hat{m} \right) \right]. \end{aligned} \quad (56)$$

We can attempt to solve this equation as in the annealed case by setting derivatives with respect to  $m$ ,  $\hat{m}$ ,  $q$  and  $\hat{q}$  to zero. These equations can be written

$$\hat{m} = \alpha \beta \frac{(2\kappa - \kappa^2)m}{1 + \beta(1-q)} \quad (57)$$

$$m = \int dz \tanh(\sqrt{\hat{q}}z + \hat{m}) \quad (58)$$

$$\hat{q} = \alpha \beta \frac{q - (2\kappa - \kappa^2)m^2}{[1 + \beta(1-q)]^2} \quad (59)$$

$$q = \int dz \tanh^2(\sqrt{\hat{q}}z + \hat{m}). \quad (60)$$

First we should note the solution  $m = \hat{m} = q = \hat{q} = 0$ , which is the same (paramagnetic) free energy as in the annealed and TAP analysis, and describes the high-temperature solution. The local stability of this solution at  $\kappa = 1$ , which determines the critical temperature, is identical to that determined by alternative analysis methods (Secs. VII D and III A).

Next consider the subspace with  $\hat{q} = 0$  and  $m \neq 0$ . In this case we find the relationship  $q = (2\kappa - \kappa^2)m^2$ , and  $q = 0$ , and recover the planted phase with identical properties to the annealed and TAP analysis. The energy of this solution being controlled by  $\kappa$ . This solution, where it exists is again locally stable.

Finally, for  $\alpha > 1$  a solution with nonzero  $\hat{q}$  is possible in the subspace with  $m = 0$ , the critical temperature given by  $T = -1 + \sqrt{\alpha}$ . Nevertheless, for  $\kappa = 1$  the embedding-aligned phase ( $|m| > 0$ ) is dominant. Thus the replica symmetric solution is in agreement with the understanding presented in the main text. For  $\rho < 1$  this solution can become relevant for a limited region

$\alpha \gtrsim 1$ , and describes a transition similar to that in the SK model.

However, it is known that the replica symmetric solution can be unstable, and it is necessary to go beyond these approximations to first and higher orders of replica symmetry breaking to describe the phase space correctly, this being most important at small  $\alpha$ . We have not probed the interesting consequences in this part of the phase diagram, but the results of the anti-Hopfield model are understood to apply and there is expected to be dynamical and static transitions in the space orthogonal to the planted solution [36, 69]. It is the existence of these solutions, whether they are stable or metastable, alongside the planted one, that offers the interesting possibility to undertake a *population transfer* measurement in the context of this model [27]. Being able to plant a deep stable or metastable solution may also have other interesting applications, particularly in sampling and inference applications where one must provide more than a single ground state certificate.

The aim of this paper is to provide practical intermediate scale benchmarks, the replica method describes only the typical case properties at leading order in  $N$ . A tension therefore exists between the results found in this section, and practical limitations of the ensemble such as precision, and instance to instance fluctuations at finite size. These facts must always be borne in mind when considering such analysis results.

## F. Time to solution measurements

In this work the problem hardness is quantified via time to solution (TTS) of parallel tempering Monte Carlo. We do emphasize that we expect similar results using other heuristics. For a single disorder realization, the time to solution is defined as the run time such that there is a

99% success probability to have found the solution at the end of the run.

Many solvers have parameters that affect the success probability which we denote by a set  $\{\phi\}$ . For parallel tempering Monte Carlo, the parameters considered are the lowest temperature and number of replicas. Often, it is faster to carry out many short attempts taking run time  $R$  with a lower success probability  $p(R, \{\phi\})$  which motivates the definition [70–72]

$$\text{TTS}(R, \{\phi\}) \triangleq R \frac{\log(1 - 0.99)}{\log(1 - p(R, \{\phi\}))}. \quad (61)$$

The number of attempts is taken to be a real number for convenience.

In the case of an ensemble, averages may be poorly defined and so the median is used [70]. For the TTS to be well defined, the minimum value with respect to run time and parameters must be computed [71]. The ensemble TTS is defined as

$$\text{TTS} \triangleq \min_{R, \{\phi\}} \text{median}_i \{\text{TTS}_i(R, \{\phi\})\},$$

where the index  $i$  refers to the TTS of individual disorder samples as a function of run time and parameters.

Due to the nonsequential nature of parallel tempering Monte Carlo, the success probability as a function of run time can be efficiently measured with only a small number of runs [73]: The algorithm is run  $W$  times until the solution is found. For a given run time  $R$ , the success probability is estimated as the percentage of runs where the solution is found in time less than  $R$ .

Due to finite numerical precision, solutions are considered to be any state with energy  $E < E_{\text{GS}} + \epsilon$  where  $E_{\text{GS}}$  is the planted solution energy and  $\epsilon$  is a numerical constant.  $\epsilon = 10^{-7}$  was used for the TTS study in Fig. 10. The effects of different values of  $\epsilon$  on the location of the TTS maximum are shown in Fig. 14.

- 
- [1] T. Hogg, B. A. Huberman, and C. P. Williams, *Phase transitions and the search problem* (Elsevier, Amsterdam, 1996).
  - [2] R. Monasson, R. Zecchina, S. Kirkpatrick, B. Selman, and L. Troyansky, *Determining computational complexity from characteristic ‘phase transitions’*, *Nature* **400**, 133 (1999).
  - [3] M. Mézard, G. Parisi, and R. Zecchina, *Analytic and Algorithmic Solution of Random Satisfiability Problems*, *Science* **297**, 812 (2002).
  - [4] L. Zdeborová and M. Mézard, *Locked Constraint Satisfaction Problems*, *Phys. Rev. Lett.* **101**, 078702 (2008).
  - [5] M. W. Johnson, M. H. S. Amin, S. Gildert, T. Lanting, F. Hamze, N. Dickson, R. Harris, A. J. Berkley, J. Johansson, P. Bunyk, et al., *Quantum annealing with manufactured spins*, *Nature* **473**, 194 (2011).
  - [6] Z. Wang, A. Marandi, K. Wen, R. L. Byer, and Y. Yamamoto, *Coherent Ising machine based on degenerate optical parametric oscillators*, *Phys. Rev. A* **88**, 063853 (2013).

- [7] W. Barthel, A. K. Hartmann, M. Leone, F. Ricci-Tersenghi, M. Weigt, and R. Zecchina, *Hiding solutions in random satisfiability problems: A statistical mechanics approach*, Phys. Rev. Lett. **88**, 188701 (2002).
- [8] F. Krzakala and L. Zdeborová, *Hiding quiet solutions in random constraint satisfaction problems*, Phys. Rev. Lett. **102**, 238701 (2009).
- [9] F. Ricci-Tersenghi, G. Semerjian, and L. Zdeborová, *Topology of phase transitions in Bayesian inference problems*, Phys. Rev. E **99**, 042109 (2019).
- [10] I. Hen, J. Job, T. Albash, T. F. Rønnow, M. Troyer, and D. Lidar, *Probing for quantum speedup in spin glass problems with planted solutions* (2015), (arXiv:quant-ph/1502.01663).
- [11] W. Wang, S. Mandrà, and H. G. Katzgraber, *Patch-planting spin-glass solution for benchmarking*, Phys. Rev. E **96**, 023312 (2017).
- [12] F. Hamze, D. C. Jacob, A. J. Ochoa, D. Perera, W. Wang, and H. G. Katzgraber, *From near to eternity: Spin-glass planting, tiling puzzles, and constraint-satisfaction problems*, Phys. Rev. E **97**, 043303 (2018).
- [13] I. Hen, *Equation Planting: A Tool for Benchmarking Ising Machines* (2019), (arxiv:quant-ph/1903.10928).
- [14] C. Pattison, F. Hame, J. Raymond, and H. G. Katzgraber, in *APS Meeting Abstracts* (2018), p. S28.004.
- [15] D. J. Thouless, P. W. Anderson, and R. G. Palmer, *Solution of 'Solvable model of a spin glass'*, Phil. Mag. **35**, 593 (1977).
- [16] C. Geyer, in *23rd Symposium on the Interface*, edited by E. M. Keramidas (Interface Foundation, Fairfax Station, VA, 1991), p. 156.
- [17] K. Hukushima and K. Nemoto, *Exchange Monte Carlo method and application to spin glass simulations*, J. Phys. Soc. Jpn. **65**, 1604 (1996).
- [18] S. Kirkpatrick, C. D. Gelatt, Jr., and M. P. Vecchi, *Optimization by simulated annealing*, Science **220**, 671 (1983).
- [19] S. F. Edwards and P. W. Anderson, *Theory of spin glasses*, J. Phys. F: Met. Phys. **5**, 965 (1975).
- [20] M. Blume, *Theory of the First-Order Magnetic Phase Change in  $\text{UO}_2$* , Phys. Rev. **141**, 517 (1966).
- [21] H. W. Capel, *On the possibility of first-order phase transitions in Ising systems of triplet ions with zero-field splitting*, Physica **32**, 966 (1966).
- [22] J. S. Yedidia and A. Georges, *The fully frustrated ising model in infinite dimensions*, J. of Phys. A: Mathematical and General **23**, 2165 (1990).
- [23] D. Sherrington and S. Kirkpatrick, *Solvable model of a spin glass*, Phys. Rev. Lett. **35**, 1792 (1975).
- [24] J. J. Hopfield, *Neural networks and physical systems with emergent collective computational abilities*, Proc. Natl. Acad. Sci. **79**, 2554 (1982).
- [25] D. J. Amit, H. Gutfreund, and H. Sompolinsky, *Statistical Mechanics of Neural Networks near Saturation*, Ann. Phys. **173** (1987).
- [26] H. Nishimori and K. Takada, *Exponential Enhancement of the Efficiency of Quantum Annealing by Non-Stoquastic Hamiltonians*, Frontiers in ICT **4**, 2 (2017).
- [27] V. N. Smelyanskiy, K. Kechedzhi, S. Boixo, S. V. Isakov, H. Neven, and B. Altshuler, *Non-ergodic delocalized states for efficient population transfer within a narrow band of the energy landscape* (2018), (arxiv:quant-ph/1802.09542).
- [28] C. H. Papadimitriou and K. Steiglitz, *Combinatorial optimization: algorithms and complexity* (Courier Corporation, 1998).
- [29] Y. T. Fu and P. Anderson, *Lectures in the Sciences of Complexity* (Addison-Wesley, Reading, MA, 1989).
- [30] S. Mertens, *A physicist's approach to number partitioning*, Theoretical Computer Science **265**, 79 (2001).
- [31] C. Borgs, J. Chayes, and B. Pittel, *Phase transition and finite-size scaling for the integer partitioning problem*, Random Structures & Algorithms **19**, 247 (2001).
- [32] N. Karmarkar, R. M. Karp, G. S. Lueker, and A. M. Odlyzko, *Probabilistic analysis of optimum partitioning*, Journal of Applied Probability **23**, 626 (1986).
- [33] J. C. Lagarias and A. M. Odlyzko, *Solving low-density subset sum problems*, Journal of the ACM (JACM) **32**, 229 (1985).
- [34] M. R. Garey and D. S. Johnson, *Computers and intractability: a guide to the theory of NP-completeness* (Freeman, San Francisco, 1979).
- [35] I. P. Gent and T. Walsh, in *ECAI* (PITMAN, 1996), p. 170.
- [36] K. Nokura, *Spin glass states of the anti-Hopfield model*, J. Phys. A: Mathematical and General **31**, 7447 (1998).
- [37] H. Uhlig, *On singular Wishart and singular multivariate beta distributions*, The Annals of Statistics p. 395 (1994).
- [38] R. M. Karp, *Complexity of Computer Computations* (New York: Plenum, 1972), chap. Reducibility among Combinatorial Problems, p. 85.
- [39] C. H. Papadimitriou, *On the complexity of integer programming*, Journal of the ACM (JACM) **28**, 765 (1981).
- [40] R. Merkle and M. Hellman, *Hiding information and signatures in trapdoor knapsacks*, IEEE transactions on Information Theory **24**, 525 (1978).
- [41] A. K. Lenstra, H. W. Lenstra, and L. Lovász, *Factoring polynomials with rational coefficients*, Mathematische Annalen **261**, 515 (1982).
- [42] S. Mertens, *The easiest hard problem: Number partitioning*, Computational Complexity and Statistical Physics **125**, 125 (2006).
- [43] H. Nishimori, *Statistical Physics of Spin Glasses and Information Processing: An Introduction* (Oxford University Press, New York, 2001).
- [44] G. Parisi, *Infinite number of order parameters for spin-glasses*, Phys. Rev. Lett. **43**, 1754 (1979).
- [45] M. Oppen and O. Winther, *From naive mean field theory to the TAP equations*, Advanced mean field methods: theory and practice pp. 7–20 (2001).

- [46] T. Plefka, *Convergence condition of the TAP equation for the infinite-ranged Ising spin glass model*, J. Phys. A: Mathematical and General **15**, 1971 (1982).
- [47] T. Tanaka, *Information geometry of mean-field approximation*, Neural Computation **12**, 1951 (2000).
- [48] M. Mézard, G. Parisi, and M. A. Virasoro, *Spin Glass Theory and Beyond* (World Scientific, Singapore, 1987).
- [49] D. J. Amit, H. Gutfreund, and H. Sompolinsky, *Spin-glass models of neural networks*, Phys. Rev. A **32**, 1007 (1985).
- [50] K. Nakanishi and H. Takayama, *Mean-field theory for a spin-glass model of neural networks: TAP free energy and the paramagnetic to spin-glass transition*, J. Phys. A: Mathematical and General **30**, 8085 (1997).
- [51] M. Shamir and H. Sompolinsky, *Thouless-Anderson-Palmer equations for neural networks*, Phys. Rev. E **61**, 1839 (2000).
- [52] G. Parisi and M. Potters, *Mean-field equations for spin models with orthogonal interaction matrices*, J. Phys. A: Mathematical and General **28**, 5267 (1995).
- [53] V. A. Marchenko and L. A. Pastur, *Distribution of eigenvalues for some sets of random matrices*, Matematicheskii Sbornik **114**, 507 (1967).
- [54] M. Mezard and A. Montanari, *Information, physics, and computation* (Oxford University Press, 2009).
- [55] K. Binder, *Critical properties from Monte Carlo coarse graining and renormalization*, Phys. Rev. Lett. **47**, 693 (1981).
- [56] H. G. Katzgraber, M. Körner, and A. P. Young, *Universality in three-dimensional Ising spin glasses: A Monte Carlo study*, Phys. Rev. B **73**, 224432 (2006).
- [57] K. Binder, K. Vollmayr, H.-P. Deutsch, J. D. Reger, M. Scheucher, and D. P. Landau, *Monte Carlo Methods for First Order Phase Transitions: Some Recent Progress*, Int. J. Mod. Phys. C **3**, 1025 (1992).
- [58] K. Vollmayr, J. D. Reger, M. Scheucher, and K. Binder, *Finite size effects at thermally-driven first order phase transitions: A phenomenological theory of the order parameter distribution*, Zeitschrift für Phys. B Condens. Mat. **91**, 113 (1993).
- [59] J. J. Moreno, H. G. Katzgraber, and A. K. Hartmann, *Finding low-temperature states with parallel tempering, simulated annealing and simple Monte Carlo*, Int. J. Mod. Phys. C **14**, 285 (2003).
- [60] Z. Zhu, C. Fang, and H. G. Katzgraber, *borealis - A generalized global update algorithm for Boolean optimization problems* (2016), (arXiv:1605.09399).
- [61] H. H. Hoos and T. Stützle, *Stochastic local search: Foundations and applications* (Elsevier, 2004).
- [62] Indeed, due to rounding errors, it can happen that the planted solution's energy exceeds that of some other spurious state by a tiny amount!
- [63] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables* (Dover, New York, 1964).
- [64] G. H. Golub and C. F. Van Loan, *Matrix computations*, vol. 3 (JHU press, 2012).
- [65] A. J. Bray and M. A. Moore, *Metastable states in spin glasses*, J. Phys. C **13**, L469 (1980).
- [66] P. Garstecki, T. X. Hoang, and M. Cieplak, *Energy landscapes, supergraphs, and "folding funnels" in spin systems*, Phys. Rev. E **60**, 3219 (1999).
- [67] In practice however, i.e. when different optimization algorithms are concerned, this argument might not be strictly applicable as different optimization routines use different methodologies to access the possible state space of the systems.
- [68] Y. L. Tong, *The multivariate normal distribution* (Springer Science & Business Media, 2012).
- [69] G. Parisi, *On the Statistical Properties of the Large Time Zero Temperature Dynamics of the SK Model*, Fractals (Supplementary Issue) **11**, 161 (1995), (cond-mat/9501045).
- [70] S. Boixo, T. F. Rønnow, S. V. Isakov, Z. Wang, D. Wecker, D. A. Lidar, J. M. Martinis, and M. Troyer, *Evidence for quantum annealing with more than one hundred qubits*, Nat. Phys. **10**, 218 (2014).
- [71] T. F. Rønnow, Z. Wang, J. Job, S. Boixo, S. V. Isakov, D. Wecker, J. M. Martinis, D. A. Lidar, and M. Troyer, *Defining and detecting quantum speedup*, Science **345**, 420 (2014).
- [72] T. Albash and D. A. Lidar, *Demonstration of a Scaling Advantage for a Quantum Annealer over Simulated Annealing*, Phys. Rev. X **8**, 031016 (2018).
- [73] S. Mandrá and H. G. Katzgraber, *A deceptive step towards quantum speedup detection*, Quantum Sci. Technol. **3**, 04LT01 (2018).